

Result No.	Query			Length	DB	ID	Description
	Score	Match	%				
1	5180.8	98.4	14973	1	AE001314	AE001314	Chlamydia
2	2456.6	46.7	12838	1	AE002338	AE002338	Chlamydia
3	824.2	15.7	10199	1	AE001315	AE001315	Chlamydia
4	435.8	8.3	5172	16	AB033817	Ab033817	Chlamydia
5	435.8	8.3	14781	1	AE001638	AE001638	Chlamydia
6	435.8	8.3	20517	1	AE002181	AE002181	Chlamydia
7	435.8	8.3	300550	2	AP002547	AP002547	Chlamydia
8	61.8	1.2	117655	75	AC074221	AC074221	Mus. musculus

9	61	1.2	1159	89	AF126748	Homo sapi	AF126748 Homo sapi
c 10	61	1.2	1442	88	AF087653	Homo sapi	AF087653 Homo sapi
c 11	61	1.2	1472	88	AF126749	Homo sapi	AF126749 Homo sapi
12	60.4	1.1	720	85	AB007820	Homo sapi	AB007820 Homo sapi
13	60.4	1.1	1410	85	AB009843	Homo sapi	AB009843 Homo sapi
14	60.4	1.1	2479	95	RNU87960	Rattus norv	U87960 Rattus norv
c 15	58.8	1.1	624	5	AF139019	Cepaea ne	AF139019 Cepaea ne
16	57.8	1.1	222172	75	AC079245	Mus muscu	AC079245 Mus muscu
17	57.6	1.1	471	97	HSU23863	Human clone	U23863 Human clone
c 18	57.6	1.1	1133	53	CNS062LW	AL422490 T3 end of	AL422490 T3 end of
19	57.6	1.1	3103	97	HSU80742	U80742 Homo sapien	U80742 Homo sapien
20	57.6	1.1	4751	91	BC004354	Homo sapi	BC004354 Homo sapi
21	57.6	1.1	5120	10	AX067148	Sequence	AX067148 Sequence
22	57.6	1.1	5120	89	AF152102	Homo sapi	AF152102 Homo sapi
23	57.6	1.1	6604	91	D83783	Human mRNA	D83783 Human mRNA
24	57.6	1.1	6757	89	AF117755	Homo sapi	AF117755 Homo sapi
c 25	57.6	1.1	6788	88	AF071309	Homo sapi	AF071309 Homo sapi
26	57.4	1.1	186838	67	AC032286	Mus muscu	AC032286 Mus muscu
27	57.2	1.1	195	9	AR002263	Sequence	AR002263 Sequence
28	57.2	1.1	195	9	AR053097	Sequence	AR053097 Sequence
29	57.2	1.1	234	9	AR002264	Sequence	AR002264 Sequence
30	57.2	1.1	234	9	AR053098	Sequence	AR053098 Sequence
31	57	1.1	299	92	HSA131896	Homo sapi	AU131896 Homo sapi
32	56.8	1.1	323	8	SSA012206	Salmo sal	AJ012206 Salmo sal
33	56.6	1.1	154	9	AR002267	Sequence	AR002267 Sequence
34	56.6	1.1	154	9	AR053101	Sequence	AR053101 Sequence
35	56.6	1.1	168	9	AR002265	Sequence	AR002265 Sequence
36	56.6	1.1	168	9	AR053099	Sequence	AR053099 Sequence
37	56.6	1.1	171	9	AR002266	Sequence	AR002266 Sequence
38	56.6	1.1	171	9	AR053100	Sequence	AR053100 Sequence
39	56.6	1.1	477	10	E34519	gene a	E34519 SCA7 gene a
40	56.6	1.1	477	88	AF020275	Homo sapi	AF020275 Homo sapi
41	56.6	1.1	197897	82	AP000779	Homo sapi	AP000779 Homo sapi
c 42	56.6	1.1	215963	71	AC040982	Mus muscu	AC040982 Mus muscu
43	56.2	1.1	326419	76	AC079635	Sequence	AC079635 Mus muscu
44	56	1.1	172	97	S82497	SCA1=spinoc	S82497 SCA1=spinoc
45	56	1.1	471	93	HSCA40	H.sapiens m	X85325 H.sapiens m

ALIGNMENTS

RESULT	1
AE001314	AE001314
LOCUS	AE001314 14973 bp DNA BCT 30-OCT-2000
DEFINITION	Chlamydia trachomatis section 41 of 87 of the complete genome.
ACCESSION	AE001314 AE001273
VERSION	AE001314.1 GI:3328833
KEYWORDS	Chlamydia trachomatis.
SOURCE	Chlamydia trachomatis.
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE	1 (bases 1 to 14973)
AUTHORS	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.
TITLE	Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
JOURNAL	Science 282 (5389), 754-759 (1998)
MEDLINE	99000809
PUBMED	9784136
REFERENCE	2 (bases 1 to 14973)
AUTHORS	Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE	Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL	Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE	99206606
PUBMED	10192388
REFERENCE	3 (bases 1 to 14973)
AUTHORS	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.
TITLE	Direct Submission

JOURNAL	Submitted (20-MAY-1998) Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES	Location/Qualifiers
source	1..14973
gene	/organism="Chlamydia trachomatis"
CDS	/strain="D/UW-3/CX"
	/db_xref="taxon:813"
	228..692
	/gene="CT406"
	228..692
	/gene="CT406"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAC68003.1"
	/db_xref="GI:3328834"
	/translation="MLCPFNHGELKVIDSRNAPESNAIKRRRCLRCRSORFTTFETV
	ELFVQLKRDGRVENFQESKLVAASHSHTRIGQEOVAIASINIKODLLGKONREI
	STKEIGELVMKYLKADMIAYIRFACVYRRFKDVGELMEVLLSATPDGEQT"
	706..1080
	/gene="dksa"
	/note="CT407"
	706..1080
	/gene="dksa"
	/codon_start=1
	/transl_table=11
	/product="DnaK Suppressor"
	/protein_id="AAC68004.1"
	/db_xref="GI:3328835"
	/translation="MPLTDEEIANFKTLLEMAKLSHTLEGNAQEVKKPNEATGYSO
	HQADQGTDTFTDISLEVTWKYKLLRQIDRALEKIEEASYGICDVSGETPLARLMA
	IPYATMVKSQEKFEKGLSGN"
	1086..1589
	/gene="lspa"
	/note="CT408"
	1086..1589
	/gene="lspa"
	/codon_start=1
	/transl_table=11
	/product="Lipoprotein Signal Peptidase"
	/protein_id="AAC68005.1"
	/db_xref="GI:3328836"
	/translation="MPNRSLSPTFTLLLLASIDWVKLVLLKSCQLSPHSSAFLYSY
	VWGFSLIIPSENEGAAGLFTQYKIPLIIFRVCTILGLALFRIKYKSLHRRTRVA
	LTLTAGALGNVGDIILYGVDFSLSLSYSWRPFPSNLADAFISIGTLLLIGHLYFT
	KESKYP"
	1691..3052
	/gene="CT409"
	1691..3052
	/gene="CT409"
	/codon_start=1
	/transl_table=11
	/product="Amino acid permease"
	/protein_id="AAC68006.1"
	/db_xref="GI:3328837"
	/translation="MSAFSLSPLTAFDDFIWSYIAFVILLGLVLTCKSKPAQFTQLP
	AFKLIYFHSEKSKKNGKGVPLKVFASAGNIGNVGIVTAASVGGPGAL
	FWWIAIGLSIVKYSVYLGIKFRQSDANNVHGGPMFFLDKXAYRIVSVIYAVLL
	CITGYEIQSIANTLASCWDVPLMAIASLLFLVMFVAVGGGLQRIKICSLVLPFF
	MLVYGMAFYILIOEIHITLPALFSSIFRSFTGHTAGGAGCTVASTIRGLSRAAY
	SGDIGIGDSIIQSETATNPQAOISIVGVINDVLNCTLSLLIVLASGVWHRAGLE
	SGEIVEQALSFPYIRVFLPAFLFATGYTTIISYFLVGVKKCANFICGNKSKFTYV
	GIIALPAFCFLPQDTALLVMSVGALLLCLNLLGVFLMRKELIFPKKEAETLEVSEST
	LSS"
	3267..4544
	/gene="pcnB_1"
	/note="CT410"
	3267..4544
	/gene="pcnB_1"
	/codon_start=1
	/transl_table=11
	/product="PolyA Polymerase"
	/protein_id="AAC68007.1"









TTGQVAGGAIFSPSVITITPVTFKSNAINATTSSKKTDFGGAIGAISTVLSKGFAR  
SENTADLSAIGLVPTTQDAETVQLTGTSTYFERNKALKRATVYAPIVSIKAHTATFD  
ONISAEESAIYFKEATIESLGSVFTNLVPIQSTVLTSGNTSVYGAIAFGOIA  
NASSOTDNLPKLIASGNTSFRNNEYPDATNGQSTFCSAGDIKLTWQAAEGKV  
ISFPDAJTSFKKGTGLASAYDTLDINKNSDGSINSAPTGTIMFSSHELNSKIPO  
NVVLHSGSLIKANTELHVLVSFDQREGSLIMEPGSVLSNODIAGSLVNSLITDLS  
SVGRNASGDNIFMPELURIWDTSTNSGNSSTPPSNTPNSTPTAQAIPSKNPAFT  
VLTPTPTPTTGNIFPLNGVIKILDPNGFTFQNPALGSDQKISLVLPSDQKLOAKV  
VLTDISPKGYTGTLDLPQOLQNGVIAQWTFKSYRQWAYIPRDNHFIYANSILGSO  
MSMATVKNLNDKLNLAIRDEAVYNNVLISGLGTMLSORGORSEEMTYYSRGA  
LDAPKTDOLLIAGASFKMIGRSKSLKERNYTHKGSSEYXOASVYGGSPFVLTINKEA  
GRSLPDLLOGLVSYGVIKHDVTHYPTIRELNKGEWDLGMLTALRVSSILKTPKQGD  
SKRITVYGEVETSIROKQFETEDYPRFNCTYRNLAIVPGLALEGEFKNGLIMY  
NRFVSYPSTIRNPSVPOKYQVLSSGEIGEIVCGVPTRNSSRAEYSTQLYGLPLWLT  
GSYTLAADAHTLANNINGGARMTF"  
BASE COUNT 4074 a 2917 c 2313 g 3534 t  
ORIGIN

Query Match 46.7%; Score 2456.6; DB 1; Length 12838;  
Best Local Similarity 70.0%; Pred. No. 0;  
Matches 3717; Conservative 0; Mismatches 1239; Indels 357; Gaps 16;  
QY 7 atgaatgctgcagctactcgggtgttctgtgtctgtctccctcagttcagggttt 66  
DB 3167 ATGAAATGGCTGCTAGCTACTGCTGTGTGTGCTGTCTTCTCTCAGTCTCAGGGTTT 3226  
QY 67 tgcctccagaacctaaagaataattctctcgtgagaaactctctcctcaccact 126  
DB 3227 TGTTCCTCCGGAATCTAAAGAATTGAATTTCTCTCGCACAGGAACCTCTTCTTCTACTACT 3286  
QY 127 ttactgaacaattggagaagctggggcagaataatctgtctctgtgaacgcattcttc 186  
DB 3287 TTTACAGAAACTATTGGAGAGAACGCTACAGAAATATATGTCTCTGGTAACTCGTCATTT 3346  
QY 187 acaaaatttaccacaattctactaccgcagatacaacaact-----actgctccgtt 273  
DB 3347 ACTAATTTTCTACTAATATTCTCAGTCAAAAAACCAACGACTGATGACTCAAGTAGCTCAACC 3406  
QY 227 ccacgaactcaactcctctagctctagcgagaa-----actgctccgtt 273  
DB 3407 CCTACCACATCTCTGCTGTAGATCAACAGAAAGATGTGTCAGGCCCTCTCTCTCATCT 3466  
QY 274 tctgaggaagtagtactctacaacagactcctgatacctaaagggtgcgccgtttat 333  
DB 3467 TCGCCTAATCTGGGATACGTACGTACTCTCCGATCCTTAAAGGTGGCGAGCTTTTAC 3526  
QY 334 aacgcgcactcggagtttcttcttatgacacgatacaggaacagaaggttctcctaact 393  
DB 3527 AATGAACACTCCGGAAATTTATCTTTATGGCAAGATCTGGAGTAGAAGGTTCTCTCACT 3586  
QY 394 ctgctcagagataaaatgactgtgaaggcgtgcttatctctcctcaaggagagctgcta 453  
DB 3587 CTATCTAATATCAAGTAGACCGGTGATGGAGAGCTTATCTATTTCCCAAGCGCAACTCTTA 3646  
QY 454 ttacagactctgacagcttaaccatcccaataacttatccagactatcccgagagagcg 513  
DB 3647 TTTACAGATCTAATCTGTTTAACTTCAAGGAATCTATCCCAATTAATCTGGAGAGGA 3706  
QY 514 atttttgaggagctacatactcccttatcagaggtattactaaagcagacttctcctgcaac 573  
DB 3707 ATCTTTTGGAGGCTCTACTATTTCTCTGGAATCAATCAAGCCACATTTCTCATCAAT 3766  
QY 574 tctgcagagactcctgctcctgttaagaacctacagaaacctaaagctcaaacgacaac 633  
DB 3767 ACTGCAGAAAGTT-----GTCCCTGAAGAACTACACCGAAC 3802  
QY 634 gaaacgtcgggtcttagtattctagtcggaatgattcgggttcttccccccaggttccagt 693  
DB 3803 CCAATCCAGGAACCTCAACAACACACCTCTCACACCATCACCTCC----- 3850  
QY 694 agagctgaacccgcagcagctaattctcaagctcaacttatttctgctacagctactact 753

DB 3851 -----AAGTCCAAAGCTATTATTCATTACTTCTTCTACTACTCA 3889  
QY 754 gctgcctcaaacccagatacaaaaacatcaactccctctcataagccagatctcggggaect 813  
DB 3890 GCTAATGGAACCGGTCCAGATTCTCAAACTCCATCCCAATAAACCCGGATCTGGAGGACT 3949  
QY 814 atctatgtctaaaggcgaccttactatcgcagactctcaagagctactattctcctaataat 873  
DB 3950 ATCTATGCCACAGGAGATTTAACCAATTTCCGACTCTCAAGAGATCGTTTTCTCTCCGTAAAC 4009  
QY 874 aaagctactaaagatgagagagcgatctcttctgctgaagaagatgttctcttcgagaataatt 933  
DB 4010 AAAGCTAGTAAAGATGAGGAGGCCATCTTTGCAAAAATAATGTTCTTCTCGAAGATATC 4069  
QY 934 acatcattaaaagtacaacaactaacggtgctggaagaaggagagctactctatgctaaa 993  
DB 4070 ACNACACTAAAGTACAAAACAATGGAGCAGAGGAGGAGGCGGTATCTATGCTAGT 4129  
QY 994 ggtgacctctcaattcaattctttaaacaagagctctttaaattcttaactacagtaaacaa 1053  
DB 4130 GGAGATCTCTCCATTTCAGTCTCTAAACAAGAGCTTTTAAATTTCCAATACCAGCAAGCA 4189  
QY 1054 ggtgggggggctctatatgttgaaggaggtataaaacttccaagatcttgaagaataattcgc 1113  
DB 4190 GGAGCGGAGCTCTTTATATTGAAGGGAATGTAGACTTCAAGATCTCGAAGAAATTCGC 4249  
QY 1114 attaatcaataaagctggaacgtttcgaacaaacaaacaaacaaacaaacaaacaaacaaac 1173  
DB 4250 ATTAAGTACATAAATCAGGAACCTTTCGAAACAAAGAAAGTTACTTTATGCTTACCCTGAA 4309  
QY 1174 gctcaagcatctgcaggaaatgcagatgctgtgggctctctcctcctcctcaatcgtgtct 1233  
DB 4310 GGGCA----- 4314  
QY 1234 ggaacactacagctctccagactcagagagactctagctctggtcagactcggataactca 1293  
DB 4315 --GACAAAATAAAGCTCTGTAACTGTCTTCCAGTCTCCAGTCCCAATAATCTACACCACT 4372  
QY 1294 gaaacgttccagctcagactaaagcggtgggcttatactgataagaactcttcgatt 1353  
DB 4373 CCTACTCTCCAGTCAAGCTTAAAGGTGGAGGTTTATACCCGAAAAAATCTTTCCTCCTC 4432  
QY 1354 actaacatcacaggaaattatcgaaattgcaaaatacaaacagcgacagatgctggaggtt 1413  
DB 4433 TCTAATATTACAGGGATTTATGAAATTTACTAACAATAAAGCAACGAGTGTGGAGGAGGA 4492  
QY 1414 gcttaagtaaaagggaaccttactgtgaacactctcacgctctacaaatttttgaanaac 1473  
DB 4493 GCTTACGTAAAAAGGAACACTCACTTGCAAAAGATTCTCATCGCTTACAAATTTTCAGAAAAC 4552  
QY 1474 tcttccgataaaaggtgaggaatctcagagagagacaacatcacctctatctaatctg 1533  
DB 4553 TCTTCCGAGAAAAGAGGAGGAGCTTTACACAGAGATACGATTTACCTTATCTAACCTG 4612  
QY 1534 acaggaagactctatttccaaagagaaatactgcaagaagagagcggtgagactcttcaata 1593  
DB 4613 ACAGGTAAACCTCTATTTCGAAGAGATACCGCTTAAAGAGAGAGTGGGGCTCTTTACATC 4672  
QY 1594 aaaggtacagataaagctcttcaaatgacagagctggaatagttctctgttaataataaac 1653  
DB 4673 CAAGGAGATGACAAAACCTTTAACCATGACAGGACTCGATAGCTTCTGCTGATCGATAAC 4732  
QY 1654 acatcagaaaaacatggtgtgagcctttgttaccagaagaatactcagactctacac 1713  
DB 4733 ACTTCAGCAACCCATGGTGGAGCCTTACGTACCTTAAAGAGATTTCTCAAACTTATCT 4792  
QY 1714 tctgagtgtgaacaacttccaggaaatacagcctgtgactggtgaacagcttacttactgct 1773  
DB 4793 TCCGAGTTGAAGAGTTCCTGGAATCACTCCTGCTGCTGATGGAGAAACATCATCTCGG 4852  
QY 1774 aataaactcacaggaggttaattggtgagcggtgtgtacaaaaacgtcttgccttatcaac 1833  
DB 4853 AATAAAGCTACTGTTGTTAGTGTGGATGTATGCACAAAGCATCTTGTGTTTGTCCAAAT 4912



Db 6887 G---TTGCGAGCTACAAATCCACCACTGGCAGACAAAGAGTCTTCTTAACCGCGCTCTG 6943  
QY actttaataagatcctaataagaaacttttaccaaaacccctatgttaggaagcgcatcta--- 4092  
Db 6944 ACTTTGATCGATCCAGATGGAACCTTTTACCAAAATCCTATCTTAGGGACAGATTTAAC 7003  
QY gatgtaccactaataaagcttcocgactaacacaagaagtcagctcgaagtcctatgatttaact 4152  
Db 7004 GATGTACCGCTTAATAAATCCCTACACAGCAAAATCAAGTTGATGTTTCTTAACCTAAACA 7063  
QY ttatctgggagatctttccctcagaagaaggtacatggaacctggacatagattctaat 4212  
Db 7064 CTATCCGGAGACCTTTCTCTAGAAGGGGTATACAGGAACCTTGGACGCTAAATCCTGAT 7123  
QY ccacaaacagggaaacttcaagccagatggacattcgatcacctatcgctcgctgggtatatac 4272  
Db 7124 CCACAACAGGAAAGTTGTCGCAACTGGAAATTCGACATGTACCGCGTTGGGAATAC 7183  
QY atacctaggagataatcttttatcgaaactctatcttaggtctcccaaaccaatgatt 4332  
Db 7184 ATACCAAGAGATAACCATTTCTATGCAATTCGATCTTGGGCTCCCAAACTCCCATGATT 7243  
QY gttgtgaagcaaggcttatcaacaacatgttgaataatgcccgcttcgatgatcgcgt 4392  
Db 7244 GTTGTGAACAAGGGCTTATTAATAACATGTTACATAATGCCCGTTTCGATGACGCTGT 7303  
QY tacaataactctgggttcaggagtaggaactttcttagctcaacaagaactcctct 4452  
Db 7304 TACACAACTCTGGGCTCAGAGTTGGAACCTTCCATTCGCAACAAAGTACTCCTCTT 7363  
QY tccgaagaattcagttactacagcgcggaacttcagttggcatcgatgccaaacctaga 4512  
Db 7364 TCTGAAGAGTTAGTTATTATAGCCGAGGAACCTCTGTAGCTATCGATGCAAAACCCACGA 7423  
QY caagattttatcctaggagctgattagtaagatagtagtgagggaagaaacccatcaaa 4572  
Db 7424 CCAGATTTTATCCTGGGAGCTGCTTTAGTAAGATGGTAGGAAGAACTAAAGCGATTAG 7483  
QY aaaaatgataaataacttcatactcaaggctctgagtagctcttaccagaactctctctatgga 4632  
Db 7484 AAGGTACATAAATTATCCCATAGGGATCCGAATACTCATACCAACGCTCTGTTTACGGA 7543  
QY gataaattcctgattcttctgtcctaataagcaacatagttgtgggcaacttcttctcctaata 4692  
Db 7544 GGTAAAGTTCTGCTACTTCTTACTTAATAAACAACATGGCTGGGCACACTACCTTCTTTATTA 7603  
QY caaggagctgctcctatggacatatataaacatgatacaacaacatttacccttctatc 4752  
Db 7604 CAAGGGGTCGTATCCTATGGGCATATTAAGCATGATACAAACACCTCTATCCTTCTATC 7663  
QY catgaaagaaataaaggagatgggaagatttaggatggtagcgatcttcgtatctct 4812  
Db 7664 CATGAAAAAATAAAGGACACTGGGAAGATCTGGTGTGTTAGTAGATCTACCGCTCTCT 7723  
QY atgattcttaagaaccttctaagaattcttctaacggaatcactgtctatggggaactc 4872  
Db 7724 ATGGATGTTAAAGAACCTTCCAAGAGATCTTCTTAACCGGCTTGCTTTTACGGAGAAT 7783  
QY gagtattccagcatttcgcagaaacagtttcacagaaatcgattacagatccaaagacattc 4932  
Db 7784 GAGTACTCAGCATTCGGTCAGAAATCATTCACGGAAATTTGACTACGATCCTCGACGCTTT 7843  
QY gatgattgcttcacagaatctgctgcctctctggtggtgagtcgctgcgaagagcattc 4992  
Db 7844 GATGATTGCTTACGAAACCTATCCGATTCCTATGGGCTGCTATTTTGAAGGGGCTATC 7903  
QY atgaactgtaattattcttatataaagcttgcatagcctacatgccttcttatctac 5052  
Db 7904 ATGAGTTATGATATCTCATGTATACAAGCTTTTCCTTAGCCTATATGCACTATTTTAC 7963  
QY agaaataatcctctgctgtaaaatcggggtatgtcttcgaatgaagctgggtcaagtattc 5112

Db 7964 AGAAATAACCCCTGCTCTGTAAATATTGGGTCCCTATCTTCGAATGAAACTAGCAAAAGTTGTC 8023  
QY tgcggagtgccaaactagaacctctgtagagcaagaatacagtaactcaactatcttgg 5172  
Db 8024 TGTGGAGTCCCCACAAACATCCGCTCGAGCAGAGATATAGTACACAACCTTTATCTAGGG 8083  
QY ccccttgagactctctacggaatactatactatcgatgtaggcattgatacgctatcgcaa 5232  
Db 8084 CCTTCTGGACCTTATATGGGAATTACCATAGACGTGGTATGTACACATTTGGCACAA 8143  
QY atgactagctcggtgctcgcatgattcttctaa 5265  
Db 8144 ATGACGAGTTGGTGCCCTGATGATTTTCTAA 8176

RESULT 3  
AE001315  
LOCUS AE001315 10199 bp DNA BCT 30-OCT-2000  
DEFINITION Chlamydia trachomatis section 42 of 87 of the complete genome.  
ACCESSION AE001315 AE001273  
VERSION AE001315.1 GI:3328842  
KEYWORDS  
SOURCE Chlamydia trachomatis.  
ORGANISM Chlamydia trachomatis.  
REFERENCE 1 (bases 1 to 10199)  
AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.  
TITLE Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
JOURNAL Science 282 (5389), 754-759 (1998)  
MEDLINE 99000809  
PUBMED 9784136  
REFERENCE 2 (bases 1 to 10199)  
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C.J., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 3 (bases 1 to 10199)  
AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES  
source  
location/Qualifiers  
1..10199  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/Cx"  
/db\_xref="taxon:813"  
120..5432  
/gene="pmpc"  
/note="CT414"  
120..5432  
/gene="pmpc"  
/codon\_start=1  
/transl\_table=11  
/product="Putative outer membrane protein C"  
/protein\_id="AAC68011.1"  
/db\_xref="GI:3328843"  
/translation="MKFSATAVFAAALSSVTEASSIODIKNTDNCVSKLGVSTSOA  
FTDMLADNTEYRAADSVFYDFSTSSRLPRKHLSSSEASPTTGGVSSSSSGETDEK  
TEEDLNGGIYAREKLITISEQDSLNSQSIELHNSIFFGGEVIFDHRVAKNGGA  
IYGEKVFENIKSLLEVNIAYEKGVSIVAKERSYLENVEATSSNGEGGGGIY  
SQDMLISDCNNVHFQNAAGATAVKQCDEEMIVLAEVDSLSEDTLDSTPEEQY  
ESNGODGESPTQVSESPSTPSDDVLKGGIYTEKSLTITGTTIDFVSN  
ATDSGAGVFTKENLSCTNTNSLOFLKNSAGOHGGAYVTOTMSVNTTSESITTPPLI  
GEVISENTAKHGCGGICTNKLISLNKLTVTILTKSKESGGAIFTDLASITPTPE  
STPSSSPASTPEVVAIAKINRFFASTAKPAAPSLTAESQDTQDTSTFSDNSIDV  
STENILNVAIQNTSAKKGAIYGRKAKLSRLNNLELSGNSSQDVGGLCLTSEVFEF

AIGLSLHYSNAAKEGAIHSHKVTLSNLKSTFTFADNTVKAIVESTPEAPEEIPVPE
GEESTATDPSNTEGSSANTNLGSGDDTADTGTGDVNNESODTSDTGNABEEOLO
DSTOSNEENTLPSNIDQSNENTDESSDTEETIDDESVSSESSTPDGGAASS
GAPGSDQSIANACLAKSYAASDSSPVSNSSGSEEPVTSSSSDVTSADNDPSSSS
GAGSDSEEPPEAGSTTETLTLLIGGAIYGETVKLENFSGGIFSGNKAIDNTTGG
SSKSDSVLGFAGVAKTLFNLDSGSSRTVTFSGNTVSSQTDGQVAGAIYSPTVTTA
TPVFKSNKSTNANANTTDQRKDTFGGAIGATSAVLSGGAHLENVADLGSAGLV
PGTONTETVLESQSYFEKNALKRATIAVAVVSIKAYTATFNONSLEEGSAIYPT
KEASIESGVLFTGNLVTLTLSITTEGTPATSGDVTKYCAALFQGIASSNGSOTDN
LPLKLAGSGNICERNEYRPTSSDGTSTFCSTAGDVKLTMQAKAKTISFFDAIPT
STKKTQATAYDPLDINKSEDSSETVNSAFTGTLFSSSELHNKSYIPQNVLUHSGSL
VLKPTLHWISYFEKQSGSLMTPGSLVSNQTVADGALVINNMKTIDLSVENKNGIAE
GNITPPELRITDITTCGSGGTPTSDSESNQSDTEQNNDASNGESSSPA
VAAATSGRTNFAAATATPTTPTATPTTTTNOVILGGEIKLIDPNCGTFFQNPAIRD
QOISLLVPLTSSKMOAKIVLJGDIAPOKYGTCTLLDPDOLONGTILSLWKFPDSYR
QWAVPRDNHFIANSILGSOMLVTVKQGLLNDKMNARFEESYNNLWISGLTMLS
QVGPTSEETIYSRGASVALDAKPAHDVIVGAASFVKMIGKTKSLKRENNYTHKGSY
SYQASVGGKPFHEVINKTEKSLPLLQGVISYGIKHDTVTHYPTIRERNKGEWD
LGMWLTARVSLVTPAQGDTKRITVYGELEYSSIRQKQETETEDPRYDNCYRNL
AIPMLAFAGELSGNDILMYNRFVAYMLSIYRNSPTCKYQVLSGEGGEIICGVPTR
NSARGEYSTOLYLGPLWTLXGTYIEADAHTLAHMMNCGARMTF"
complement(5593..5676)
/gene="trnAser\_3"
/product="tRNA-Ser"
complement(5593..5676)
/gene="trnAser\_3"
5985..6815
/gene="yebl"
/note="CT415"
5985..6815
/gene="yebl"
/codon\_start=1
/transl\_table=11
/product="Solute-binding protein"
/protein\_id="AAC68012.1"
/db\_xref="GI:3328844"
/translacion="MRLFLILLFSLGITSYGDVSTRKQILVSIVPKFLVEQIAGD
TCQVTSVMNDHPDHPNVELSPYIEKIQVELWFKIGEGFECTERIIISCKQVDLAAN
IDKFTNGACORFLSDPTHWLSPKNLKIQIOALTEALVETAPHEHTYLRKNSLLOS
OLDLLDKISIVSSTQRWLVTHGAFAYFCRDYGFIOHTIERANSELSLSPKDVVRV
ERTIRDNHLSVILLKHAGKRSSAALVRKFNMTPIILLDPYAEDEVFNLLALATAFANL
"
6812..7522
/gene="CT416"
6812..7522
/gene="CT416"
/codon\_start=1
/transl\_table=11
/product="ABC Transporter ATPase"
/protein\_id="AAC68013.1"
/db\_xref="GI:3328845"
/translacion="MTKQMLLENVSFRYKGTGPLIVDHVSCEVYSGDFIGIIPNGGG
KTTTLQMLGLQPCVSGISITYSVQDHRPLUSIGWVPOHFSYDAAPFIVTKVELUSGRU
ATLFWYGRYQTEDEHAEALLVDLIDYKDSFSLHSGGQIQORVLLARALAPPEFL
LLDEPTANIDPVNQOKTILQILSALNKHCTILMITHDLHHTAGCFNRFVFNKTLTTILA
DTTTTISERFCNTEGRCp"
7513..8394
/gene="CT417"
7513..8394
/gene="CT417"
/codon\_start=1
/transl\_table=11
/product="(Metal Transport Protein)"
/protein\_id="AAC68014.1"
/db\_xref="GI:3328846"
/translacion="MPMISILCSLFPPLPFLSLAAGFASIAAGIIGSYIVVKRIVSI
SGWSTHSLGIGVIGALWQFNLPISPLHGAIASAIFFAIVCIGNVLKHYERDSII
SMIWSIGMAIGICISKLPSFNSLSPLFGNLIWVTPQDLYFLGLDLDFLTVAVTSC
HTRFLALCFDEKYNALHYSIKTWYLLLLILTAITTVVLMYVMGVILMLSMLVPSVI
ACRSYKMSHIIYIASILNIVCSFLGIMLAYLLDLPVGPVAILMGGAYSLSLLKRS
YNASTPSPVSPESKINS"
complement(8310..9317)
/gene="yhb2"

/note="CT418"
complement(8310..9317)
/gene="yhb2"
/codon\_start=1
/transl\_table=11
/product="GTP Binding Protein"
/protein\_id="AAC68015.1"
/db\_xref="GI:3328847"
/translacion="MFVDOITLLELRACKNGNVVWKRKEYLPKGGPYGGNGGSGSI
LIRATWNYSFEEYRNLRLKADGQAGASNNRTGRNGKDLVLKVPSPGTLIRDNATGE
LIHDTFKDGERIVVCOGRGKGNVFKFTSTNRAPTKATPKGPEIHLVELEKLAD
IGLVFPGNAGKSTLFTLARTVKVGAYPEFTTLHPSGLVHQEGMLQKPMWADIPG
IIEGASNRGLGDLFIERTLRLLFLFVIDISGIERHSPQDLKILMAGELLAYKEELK
DKDMVIALNKIDQLLPDEREERVALLKQFPDQFEILLGLTGBGVADLYDLFKSKLS
E"
complement(9407..9658)
/gene="rl27"
/note="CT419"
complement(9407..9658)
/gene="rl27"
/codon\_start=1
/transl\_table=11
/product="L27 Ribosomal Protein"
/protein\_id="AAC68016.1"
/db\_xref="GI:3328848"
/translacion="MAHKKGQASRNGRDSKRLGLKVGAGQVRVSTGSIILVRQGTG
WHPAVNVGRGKDDTLFALVDGIVVMKKTDRTYYSVIPQA"
complement(9689..10012)
/gene="rl21"
/note="CT420"
complement(9689..10012)
/gene="rl21"
/codon\_start=1
/transl\_table=11
/product="L21 Ribosomal Protein"
/protein\_id="AAC68017.1"
/db\_xref="GI:3328849"
/translacion="MEPYAVIOTGNKOYQVRKGDVIDVELLDGISEENKEVLFQDVLV
TFQGEKASVGAPTVGNVAVKGLSVFYRGEKVWAYKYKKRKNYHKIGHRONLYLRVKI
SDLVW"
BASE COUNT 3079 a 2313 c 1934 g 2873 t
ORIGIN
Query Match 15.7%; Score 824.2; DB 1; Length 10199;
Best Local Similarity 59.0%; Pred. No. 2.8e-166;
Matches 1765; Conservative 0; Mismatches 1018; Indels 210; Gaps 12;
Qy 2426 cagcttcagcacctgtcgcagctgctgtccctcaagcagccgcagcgcaccat 2485
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2497 CAGACTCTTCCTCATCTCGAGATAGCGCTGGAGACTCTGAAGAACCGACTGAGCCAGAAG 2556
Qy 2486 catctccagcaacacccaacttattcaggtgtagtagggaggctatctatcgagagaaaagg 2545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2557 CTGGTTCTACAAACAGAAACTCTTACTTTAATAGGAGGAGGTCTATCTATGAGAAACTG 2616
Qy 2546 ttacattctctcaatgtagcgggaactgttcagttctctcgggaacccaagctatcataaca 2605
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2617 TTAGATTGAGAACTTCTCTGCCCAAGGAATATTTCTGGAAACAAAGCTATCGGATAACA 2676
Qy 2606 -----atccctcccaatcatcgttgaaacgtacaagggaggagccatctatgccaaaacct 2659
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2677 CCACAGAAGGCTCCTCTTCCAAATCTGACGTCCTCGAGGTGCGGTCTATGCTATAACAT 2736
Qy 2660 ctttgtctattggatctctccagatgctggaacctctatattttctcggggaacagtgct 2719
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2737 TGTTTAATCTCGATAGCGGGAGCTCTAGACGAACTGTACACCTTCTCCGGAATACTGCT 2796
Qy 2720 ccaactgggaaatctcaacacacagggcaaatagcgggagcgagctactccctactg 2779
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2797 CTTC-----CAATCTACACAGAGCTCAGGTTCGTGGAGGAGCTATCTACTCTCTACTG 2850
Qy 2780 ttacattgaattgtcctcgcgacattctcttaacaatacacagcctctatagctacaccgaaga 2839

Db 2851 TAACCAATGCTACTCTCTGTAGTATTTCTTAAACAACTCTCAACAAACAATGCTAATAACA 2910  
Qy 2840 cttctctgaagatggtatcctcaggaaattctattaaagataccattgagagagccattg 2899  
Db 2911 CT-----ACAGATATCAGAGAAAGACACCTTTGGAGAGCTATCG 2952  
Qy 2900 caggga---cagccattaccctatctgaggtctctcgattttcagggatacggctgatt 2956  
Db 2953 GAGCTACTTCTGCTGTTCTCTATCAGGAGGGCTCATTTCTTAGAAAACGTTGCTGACC 3012  
Qy 2957 taggaactgaataggaactctagctaaatgcgaatacacccagtgcaactagcggatctc 3016  
Db 3013 TCGGATCTGCTATTGGTTGGTCCAGGCACACA----- 3046  
Qy 3017 aaaaatgacattacagaaaaaattactttgaaaaacggttctttatttttggaaaaaac 3076  
Db 3047 -----AATACAGAAAACAGTGAATTAGAGTCTGGCTCTACTACTTTGAAAAAATA 3099  
Qy 3077 aagctaataaacgtggagcgttactctctctagcgtttccattaaaggaataatatta 3136  
Db 3100 AAGCTTTAAACAGAGCTACTATTACGCACCTGCTCGTTTCCATTAAAGCCTATACTGCGA 3159  
Qy 3137 ccttcaatacaaatatacatccactcatgatggaaagcgtatctactttcaaaaagatgcta 3196  
Db 3160 CATTTAACCAAAACAGATCTTAGAAGAGGAAGACGCGATTACTTTTACAAAGAAAGCAT 3219  
Qy 3197 cgaattgacttttagtatgcttcttttccaggaataaacgtttacagctacacaaagcta 3256  
Db 3220 CTATTGAGTCTTTAGGCTCTGTCTCTTACAGGAACCTTAGTAACCCCTAACGCTAAGCA 3279  
Qy 3257 gttctgcaacatctggcaaaatacaaa-----atactgcaactatggggcagcca 3307  
Db 3280 CAATACAGAAGGCACACCGCCACACACCTCAGGAGATGTAAACAAATATGGTGTGCTA 3339  
Qy 3308 tctttgagatccaggaaacactcaatctctcaaacagatgccattt---taacccttc 3364  
Db 3340 TCTTTGGCAAAATAGCAAGCTCAACCGGATCTCAGACGGATACCTTCCCTCGAAACTCA 3399  
Qy 3365 ttgcttctctggaacattacttttagcaacaacagttttacagaataaccaaggtgata 3424  
Db 3400 TTGCTTCAGGAGGAATATTGTTTCCGAACAAATGAATACCGTCTCTACTTCTTGATA 3459  
Qy 3425 ctcccgctagcaagttttgattatgtagcaggatacgtcaaaactctctacagcgccta 3484  
Db 3460 CGGAAACCTCTACTTCTCTAGTATTGCGGGAGATGTTAAATTAACCATGCAAGCTGCAA 3519  
Qy 3485 aagggaagactattagcttttctgattgtgtgcacacctctaccataaaacagggttcaa 3544  
Db 3520 AAGGAAAACGATCAGTCTTTGTGATGCAATCCGGACCTCTACTAAGAAAACAGGTACAC 3579  
Qy 3545 cacaaaacgtttatgaacatttagatttaataaagaagagaa-----cagta 3592  
Db 3580 AGGCACTGCCCTACGATCTCGATATTAAATAATCTGAGGATTCAGAAACTGTAAACT 3639  
Qy 3593 atccatatacaggaaactattgttctctctctgaattatcatgaaacaactcttacatcc 3652  
Db 3640 CTGCGTTTACAGAACGATCTGTCTCTCTCTGCAATTACATGAAATAAATCTATATTC 3699  
Qy 3653 cacagaatgcaatctctcaacaggaaacttttagttctttaaagagaaaaacagaactccacg 3712  
Db 3700 CACAAAACGTAGTCTTACACAGTGGATCTTGTATTTGAAGCCCAATACCGAGCTTCATG 3759  
Qy 3713 tagtctctttgagcagaagaagggtctaaatttaattatggaacccggagctgtgttat 3772  
Db 3760 TTATTTCTTTTGGAGCAAGAAAGGCTCTCTCTCTGTTATGACACCTGGATCTGTCTTT 3819  
Qy 3773 ctaacaaacatagctaacgagctctagctatcaatgggttaaogattgatctttcca 3832  
Db 3820 CGAACGAGACTGTGTGATGGAGCTTTGGTCATAAAATAACATGACCAATTGATTTATCCA 3879  
Qy 3833 gtaagggga-----ctcctcaagcaggggaaatcttctctcctccagaaattacgtatcg 3886

Db 3880 GCGTAGAGAAAAATGTTATGCTGAAGGAAATATCTTTTACTCTCCGAATTTGAGAATCA 3939  
Qy 3887 ttgcacagacctctagtcgcatccggaggaagcgggggtcagcagtag----- 3932  
Db 3940 TAGACACTACTACAGTGGGAAGCGGTGGAACCCCATCTACAGATAGTGAAAGTAAACAGA 3999  
Qy 3933 ----- 3932  
Db 4000 ATAGTGATGATACCGAGGAGCAAAAATAAATGACGCCCTCGAATCAAGGAGAAAGCGCA 4059  
Qy 3933 -----tatacaaaaactcctaaaaagattcttctcag 3964  
Db 4060 ATGGATCGCTTCTCTCTGAGTAGTGTGTCGACACACATCTCGTCAAGAAACTTTGCGG 4119  
Qy 3965 cagtgctctcag-----gttctgcgcgaactactcctaactatgagcgagaacuaag 4015  
Db 4120 CTGCACGTACAGCCACACCTACGACAAACACCGCTACAACCTACAACAGCAACCAAG 4179  
Qy 4016 ttttctcaacagagagaccttaetttaatatagatcctaataaggaacttttaccataaacctta 4075  
Db 4180 TAATCTTAGGAGGAGAAATTAACCTCATGATCTTAATGGACCTTCTTCCAGAACCTG 4239  
Qy 4076 tgttagaagcgtatctagatgtaccactaattaagcttccgactaacacaagtgactcc 4135  
Db 4240 CATTAAGATCCGACCAACAAATCTCCTTGTAGTCTCCTACAGACTCATCAAAATGTC 4299  
Qy 4136 aagcttatgatttaacttttatctggggatcttttccctcagaaaggttacatgggaacct 4195  
Db 4300 AAGCTCAGAAAATAGTACTGACGGGTGATATGCTCTCTCAGAAAGGATATACAGGAFAC 4359  
Qy 4196 ggaacttagattctaat-----ccacaaaacagggaaacttcaagccagatggacattcgata 4252  
Db 4360 TCACTCTGGATCTGTATCAACTACAAAATGGAACGATCTCAGTGTCTTGAAATTTGACT 4419  
Qy 4253 cctatcgtcgcgtgggtatcacatcattaggaataatattttatcgaaactctatcttag 4312  
Db 4420 CTTATAGACAATGGGCTTATGTACCTTAGAGACAATCATTTCTATGCGAATCGATTTG 4479  
Qy 4313 gctcccaaaccaatgattgttggaagcaagggcttatcaacaacatgttgaataatg 4372  
Db 4480 GATCTCAATGTAAATGGTCACAGTCAACACAGGCTTGCTCAACGATAAATGAATCTAG 4539  
Qy 4373 ccgctctcgatgatatcgcttacaataaactctcgtgggtttcagggagtaggaactttcttag 4432  
Db 4540 CTGCGTTTGGAGAAGTAGCTATAACAACCTGTGGATATCAGGACTAGGAACGATGCTAT 4599  
Qy 4433 ctcaacaaggaactcctcttccgaagaattcagttactacacagccggaacttcagttg 4492  
Db 4600 CGCAAGTAGAACACCTACTTCTGGAAGAAATTCATTTATTACAGCAGGAGGCTTCTGTTG 4659  
Qy 4493 ccctcgatgccaaacctagacaagattttatcctagagctgctatttagtaagatagtg 4552  
Db 4660 CTTAGATCTAAACCCAGCCCATGATGTGTTGTTGGAGCTGCAATTTAGTAAGATGTCG 4719  
Qy 4553 ggaacacaaaagccatcaaaaaaatgcataattacttccataaagggtcctgtagtactctt 4612  
Db 4720 GGAACAAACAAATCCTTTGAAAGAGAGAAATAACTACACTCACAAGGATCGGAATATCTT 4779  
Qy 4613 accaagctctgctatggaggttaaattcctgtatttcttctcctaataagcaacatgggtt 4672  
Db 4780 ACCAAGCATCGGTATACGGAGGCCAACCATTTCCACTTTGTAATCAATAAAAAACGGAAA 4839  
Qy 4673 gggcaactctcttccataacagggagtcgtctcctatggacataataaacatgatacaaa 4732  
Db 4840 AATCGTACCGCTATTGTTACAGAGGAGTCACTCTCTACGGATATATCAACACATGATACAG 4899  
Qy 4733 caaacctttaccctctctatccattgaagaataaaggagattgggaagatttgatggatt 4792  
Db 4900 TGACTCACTATCCACGATCCGTTGAACGAAACAAAGGAGAAATGGGAAGACTTTAGGATGCG 4959  
Qy 4793 tagcggtatctgtatctctatgtatctttaaagaacacttctaagaattcttcttaaacgga 4852  
Db 4960 TGACAGCTCTCCGTGTCTCTTAAAGAACTCTGCAACAAAGGGGATACTTAAACGTA 5019











QY	4240	tggacattcgataccatctatcgctgggtatgcatacataccataggagataatcatctttttatgcg	4299
Db	12399	TGACACTTTGACAAATACCTGGCTGCCTACATCCCTAGACACAACCACTTCTACATC	12340
QY	4300	aactctcttagctcccaaaactcaaatgattgtgtgaagcaaggcttatcaacaac	4359
Db	12339	ACTCTATTGTTGGGAGCACAAACTCTTTAGTGCACCTGTGAAACAGGGATCTTAGGGAAC	12280
QY	4360	atgttgaataatgccgcttcgatgatcgcttacaataaactctgggttttcaggagta	4419
Db	12279	ATGTTGAACAATGCAAGGTTTGAAGATCGCTTTTCAACAACCTTCGGCTTCGGCTATA	12220
QY	4420	gsaactttcttagctcaacaagaactcctcttccogaagaatagttactacagccgc	4479
Db	12219	GGATCTTCTTTAGGAAGAAGATATCTGAAATTTCTGACTCATCTACCTATCATGGCAGA	12160
QY	4480	gsaactcgattgccaatgcacaaactagacaagattttctcctaagagctgcattt	4539
Db	12159	GGCTATACCGCTGCTGTGGATGCCAACCTCGCCAAAGAAATTTATTTAGGAGCTGCCTTC	12100
QY	4540	agtaagatagttggggaacaaagccatcaaaaaaatgcataacttacttccataaggcc	4599
Db	12099	AGTCAGGTTTTTGGTCACCGCAGCTGTAATATCACCTTGACAACATATAGCATAAAGGC	12040
QY	4600	tctgagtaactcttaacaagcttctgtctatgaggttaaatctctgtattcttgcctaat	4659
Db	12039	TCAGGTCACCTACACAAGCATCTCTTTATGCTGGCAATATCTTCTATTTCTCTCGGATA	11980
QY	4660	aagcaaatggttgggcaactccttctcctaatacaaggagtcgtctcctataggacattt	4719
Db	11979	C-----GGTCTCGGCTATCTTCTATCCAAAGGTGTGGCGACCTATGGTTATATG	11932
QY	4720	aaacatgatatacaacaactttaccctctctatccatgaagaataaaggagattgggaa	4779
Db	11931	CAACATGACACACACACTACTATCTCTTCTATGAGAAAAAATATGCAACCTGGGAT	11872
QY	4780	gatttaggatggttagcggtattcgtatctctatggtatcttaagaacacttctaaagat	4839
Db	11871	AGCATTTGCTTGGTTATTGATCTGGCTTTCAGTGTGGATCTTAAAGAACCTCAACCTCAC	11812
QY	4840	ttcttataacgatacactgtctatggggaactcgagatctcagacattcgcagaaaacag	4899
Db	11811	TCTACAGCAAGGCTTACCTTCTATACAGAAAGCTGAGTATACCAGAAATTCGCCAGGAGAA	11752
QY	4900	ttcacagaataatcgattacgatacgaacacttcgatgattgtgcttacagaatactgtcg	4959
Db	11751	TTACAGAGCTAGACTATGATCTAGATCTTTCTCTGCTGATGCTCTTATGGAACCTTAGCA	11692
QY	4960	cttctctgtgggatgcgctgtcgaaagagctatcatgaactgttaattattcttgtataat	5019
Db	11691	ATTCTCTAGTGGATTCTCTAGACGAGCATTAGCTTGGCGTGAGATTATTCTATATAAT	11632
QY	5020	aagcttggaatagcctatacgcctctatctacagaataaactcctgtctgttaaatcgg	5079
Db	11631	AAAGTATCAGCTGCGGTACCTCCCTGTGATCTTCAGGAATAATCCAAAACGACCTATGAA	11572
QY	5080	gtattgtcttgaatgaagctgtgaagtattctcggagattcgcgaactagaacctctgct	5139
Db	11571	GTTCTCTACAAAAGAAAAGGCAACGCTAGTCAACGTTCTCCCTACAGAAGAACGAGCT	11512
QY	5140	agagcagaataacagtaactcaatatcttctgttgcctctcttgactctctacggaaactat	5199
Db	11511	CGTGCAAGGTCAGCTCTCAAAATTTATCTTGGAAAGTTACTGGACACTCTACGGCAGTAT	11452
QY	5200	actatcgatgtaggcatgtatatacgcctatcgcaaatgactagctgcggtcgcgatgatc	5259
Db	11451	ACTATTGATGCTTCAATGAATACTTTAGTGCAAAATGGCCACGGAGGATCCGGTTTGTGA	11392
QY	5260	ttcta	5264
Db	11391	TTCTA	11387

RESULT 7  
APO02547  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AP002547 300550 bp DNA BCT 08-DEC-2000  
Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,  
section 3/4.  
AP002547 AB033786 AB033787 AB033816 AB033817 BA000008  
AP002547.2 GI:10176693  
Chlamydomophila pneumoniae J138 (strain:J138) DNA.  
Chlamydomophila pneumoniae J138  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
1 (sites)  
Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Tabuchi,M., Ouchi,K.,  
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.  
Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA  
Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
20330349  
2 (bases 1 to 300550)  
Shirai,M.  
Direct Submission  
Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.  
Mutsunori Shirai, Yamaguchi University School of Medicine,  
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,  
tel:81-836-22-2227, Fax:81-836-22-2415)  
On Sep 15, 2000 this sequence version replaced gi:6172298  
gi:6172300 gi:6172396 gi:6172398 gi:8978889.  
AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).  
Location/Qualifiers  
1. .300550  
/organism="Chlamydomophila pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
complement(21. .794)  
/gene="dapF"  
complement(21. .794)  
/gene="dapF"  
/codon\_start=1  
/transl\_table=11  
/product="diaminopimelate epimerase"  
/protein\_id="BAA98725.1"  
/db\_xref="GI:8978890"  
/translation="NAFYSPSTISKYFYISYAGNRFLLCETLPEVEDVRFQCETRDV  
GFLYKPSQSCADAQLIIFNSDGRPTMGNGLRCAIAHLASQKASDISVSDSLYS  
GYFSWDRVLVDMTLADWRASVHRLSPDRPLPKEVCILHTGPHAVVILPEISTLDL  
STLGPFLRYHOTFSPDGVNPNFVQILGHQCLRVRTYERGVETGAACGTGALASALW  
SNSYGWKESIQIHTWGGELMTVSNQRGRVYLQGSVTRDL"  
complement(763. .1338)  
/gene="clpP\_1"  
complement(763. .1338)  
/gene="clpP\_1"  
/codon\_start=1  
/transl\_table=11  
/product="CLP protease"  
/protein\_id="BAA98726.1"  
/db\_xref="GI:8978891"  
/translation="MADGEVHKLRDIIIEKLEARRVFFSEPVTEKASDAIKLWYL  
ELKDPKPIVFVINSPPGSDVAGFAVWDQIKMLTSPVTTVVTGLAASMGSVLSLCAAP  
GRRAFPHSRIMIHOPISGGPTTGATDLDIHAREILTKARIIDVYEATNQPRDII  
EKAIDRMWMTANEAKDFGLDGLFSENDL"  
complement(1358. .2917)  
/gene="glyA"  
complement(1358. .2917)  
/gene="glyA"  
/codon\_start=1  
/transl\_table=11  
/product="serine hydroxymethyltransferase"  
/protein\_id="BAA98727.1"  
/db\_xref="GI:8978892"  
/translation="MLKVFEEKFAIVEIFTKVVAVVYVLLHLKFLNENAGKKGOSLAS



Qy	4180	gggtacatgggaacctggacattagattcttaatccacaaacagggaacctccaagccaga	4239
Db	24912	GGATATTTAGGAACCTGGAAATTTGGATGCCAAATTTCCCTGGGCTTCAAAATATTTATTCTAATA	24971
Qy	4240	tggacattcgatacctatcgctcgctgggtatcacataccttagggataatacattttttatgcg	4299
Db	24972	TGGACCTTTGACAAATACCTCGCGCTGGCCCTACATCCCTAGAGACAACCACTTCTACATC	25031
Qy	4300	aacctctatttaggtccccaacatcaatgattgttgaagcaagggtcttatcaacaac	4359
Db	25032	AACCTCTATTTGGGAGAGCAAAACTCTTTAGTAGCTGTGAAACAAGGATCTTAGGGAAC	25091
Qy	4360	atgttgaataatgcccttcgatcgatcgcttacaataaactctgggttttcaggagta	4419
Db	25092	ATGTTGAACAATGCAAGTTTGAAGATCCTGCTTTCAACAACCTTCTGGGCTTCGGCTATA	25151
Qy	4420	ggaacttcttagctcaacaaaggaactcctcttccgaagaattcagttactacagccgc	4479
Db	25152	GGATCTTTCCTTAGGAAAGAAGTATCTCGAAATTCCTGACTATTCACCTATCATGGCAGA	25211
Qy	4480	ggaacttgaattgcgatcgatgccaaacctagacaagaattttacttaggagctgcattt	4539
Db	25212	GGCTATATACCGTGTGTGATGCGCAAACTCGCAAGAAATTTATTTTAGGAGCTGCCCTTC	25271
Qy	4540	agtaagatagtggggaacacaaagccatcaaaaaatgcataacttactccataaaggc	4599
Db	25272	AGTCAGGTTTTTGCTCAGCGCGATCTGAAATATCACCTTGACAACATPAAGCATAAAGGC	25331
Qy	4600	tctgagtactctaccaaagctctgtctatggaggtaaaattccgtgtatttctgtccaat	4659
Db	25332	TCAGGTCACTCTACACAAGCATCTTTATGCTGGCAATATCTCTATTTCTTCCTCCGATA	25391
Qy	4660	aagcaacatggtgggcaactccttccctaatacaagagtcgtgtcctaattggacatat	4719
Db	25392	C-----GGTCTGGCCCTATTCTATTTCCAAGGTGTGGCGACCTATGGTTATATG	25439
Qy	4720	aaacatatacaacaacattaccctctctatccatgaaagaataaaggagattggaa	4779
Db	25440	CAACATGACACCAACCTACTATCTCTATTAAGAAAGAAATATGGCAAACTGGGAT	25499
Qy	4780	gatttaggatggttagcggatcttcgtatctctatggatctttaaagaaccttctlaaagat	4839
Db	25500	AGCATTTGCTTGGTTATTGATCTGGTTTTCAGTGTGGATCTTAAAGAACCTCAACCTCAC	25559
Qy	4840	tctctaaacgatcaactgtctctatg99gaactcagtatctccagcattcgccagaaacag	4899
Db	25560	TCTACAGCAAGGCTTACCTCTATACAGAAAGCTGAGTATACAGAAATTCGCAGGAGAAA	25619
Qy	4900	tccagaaaaatcgattacgattccaaagacactcgtatgattgtgttacagaaactgtctg	4959
Db	25620	TTACACAGCTAGACTATGATCTTAGATCTTTCTCTGCATGCTCTTATGGAAACTTAGCA	25679
Qy	4960	cttccctgvggatcgctgtcgaaggagctatcatgaactgtataattctttatgtataat	5019
Db	25680	ATTCTACTGGATTTCTGTAGACGGAGCATTTAGCTTGGCGTGAGATTTATCTATATAAT	25739
Qy	5020	aagcttcattagcctacatgccttctctatcacagaataaactcgtctgtataatatcgg	5079
Db	25740	AAAGTATCAGCTGGGTACCTCCCTGTGATTTCTCAGGAATAATCCAAAAGCGACCTATGAA	25799
Qy	5080	gtaattcttcgaatgaagctggLcaagttatctgcggagtgcgaactagaaactctgct	5139
Db	25800	GTTCTCTCTACAAAAGAAAGGGCAACGTAGTCAACGTTCTTCCCTTACAAAGAACCGACGT	25859
Qy	5140	agagcagaatacacgatactcaactatacttggtccctcttgagctctctacggaactat	5199
Db	25860	CGTCAGAGGTGACCTCTCAAAATTTATCTTGGAAAGTACTTGGACACTCTACGGCACGTAT	25919
Qy	5200	actatcgatgtaggcatgtatacagctatcgcgaatgactagctgcggtgcgtcatgct	5259
Db	25920	ACTATTGATGCTTCAATGAATACATTTTATGTCGAAATGCCAAACGGAGGATTCGGGTTGTA	25979

Qy 5260 ttcta 5264  
Db 25980 TTCTA 25984

## RESULT 8

```

AC074221      AC074221      117655 bp      DNA      HTG      15-NOV-2000
LOCUS          Mus musculus chromosome 15 clone RP23-91A4 strain C57BL/6/J, WORKING
DEFINITION     DRAFT SEQUENCE, 35 unordered pieces.
ACCESSION      AC074221
VERSION         AC074221.5 GI:11177936
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE          house mouse.
ORGANISM        Mus musculus
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 117655)
Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 117655)
Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Nov 15, 2000 this sequence version replaced gi:11067106.
-----Genome Center
Center: Albert Einstein College of Medicine
Center Code: AECOM
Web site:

```

http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mousedb/mouseseqtable.htm  
Contact: jhan@sequence.aecom.yu.edu  
-----Summary Statistics-----

```

Center project name: ADZ      L08752
Sequencing vector:   puc18;
Chemistry:            Dye-terminator Big Dye;
*Consensus quality:  109029 at least Q20
*Consensus quality:  105812 at least Q30
*Consensus quality:  99196 at least Q40
Estimated insert size: agarose-pp - N/A
**Estimated insert size: 116975 - sum-of-contrigs
Quality coverage:    agarose-FP - N/A
Quality coverage:    6.2x sum-of-contrigs - N/A
=====

```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 35 contigs. The true order of the pieces is  
\* not known, and order in this sequence is arbitrary.  
\* Gaps between the contigs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence,  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	11535:	contig of 11535 bp in length
*	11536	11555:	gap of unknown length
*	11556	20397:	contig of 8842 bp in length
*	20398	20417:	gap of unknown length
*	20418	28110:	contig of 7693 bp in length
*	28111	28130:	gap of unknown length
*	28131	35775:	contig of 7645 bp in length
*	35776	35795:	gap of unknown length
*	35796	39098:	contig of 3303 bp in length
*	39099	39118:	gap of unknown length
*	39119	45564:	contig of 6446 bp in length
*	45565	45584:	gap of unknown length
*	45585	50931:	contig of 5347 bp in length
*	50932	50951:	gap of unknown length





0;

Search completed: October 26, 2001, 22:30:17  
Job time: 13750 sec

	Query Match	1.1%;	Score 58.8;	DB 5;	Length 624;
	Best Local Similarity	64.9%;	Pred. No. 0.031;		
	Matches	87;	Conservative	0;	Mismatches 47; Indels 0; Gaps 0;
QY	2394	agcaaaactctctcatccacagaggagtcgaaacacacagcttcagcaacctgctcagctgtgctgc	2453		
Db	435	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	376		
QY	2454	ttccctacaagcgccgcagcagccgcacacatcatctccagcaacacacactttaccgg	2513		
Db	375	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	316		
QY	2514	tgtagttagggaggag	2527		
Db	315	AGTAGTAGTAGTAG	302		

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 08:35:11 ; Search time 18.86 Seconds  
(without alignments)  
1912.741 Million cell updates/sec

Title: US-09-454-684A-180  
Perfect score: 8942  
Sequence: 1 MKWLSATPAVFAVLPSVSGF.....IDVGMVTLQMTSCGARMIF 1752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	374.5	4.2	1529	2	US-08-728-470-10
2	374.5	4.2	1529	4	US-08-719-641-10
3	373.5	4.2	1536	1	US-08-038-682-2
4	373.5	4.2	1536	1	US-08-302-832-2
5	373.5	4.2	1536	2	US-08-530-198-2
6	373.5	4.2	1536	2	US-08-469-880-2
7	373.5	4.2	1536	2	US-08-728-470-2
8	373.5	4.2	1536	2	US-08-617-697-2
9	373.5	4.2	1536	4	US-08-719-641-2
10	371.5	4.2	1600	2	US-08-617-697-10
11	370.5	4.1	1599	2	US-08-617-697-9
12	366.5	4.1	1338	2	US-08-728-470-9
13	366.5	4.1	1338	4	US-08-719-641-9
14	361.5	4.0	2353	4	US-09-377-155-33
15	361.5	4.0	2353	4	US-08-913-942-4
16	360	4.0	1912	1	US-08-409-995-4
17	360	4.0	1912	3	US-08-485-467-4
18	353.5	4.0	1477	1	US-08-038-682-4
19	353.5	4.0	1477	1	US-08-302-832-4
20	353.5	4.0	1477	2	US-08-530-198-4
21	353.5	4.0	1477	2	US-08-469-880-4
22	353.5	4.0	1477	2	US-08-728-470-4
23	353.5	4.0	1477	2	US-08-617-697-4
24	353.5	4.0	1477	4	US-08-719-641-4
25	352	3.9	1612	1	US-08-169-927-2
26	291	3.3	2123	4	US-08-968-685A-10
27	277	3.1	1026	1	US-08-194-290-7

28	276	3.1	1026	2	US-08-614-377A-7	Sequence 7, Appl
29	276	3.1	1026	4	US-09-142-648B-7	Sequence 7, Appl
30	272	3.0	1160	3	US-08-808-599A-24	Sequence 24, Appl
31	269	3.0	1537	1	US-08-325-267A-2	Sequence 2, Appl
32	260	2.9	1837	3	US-08-928-361B-5	Sequence 5, Appl
33	258.5	2.9	1721	3	US-08-700-651-5	Sequence 5, Appl
34	255.5	2.9	1721	3	US-08-928-361B-6	Sequence 6, Appl
35	254.5	2.8	2035	1	US-08-046-585-5	Sequence 5, Appl
36	254.5	2.8	2035	1	US-08-393-703-5	Sequence 5, Appl
37	254.5	2.8	2035	5	PCT-US93-11721-5	Sequence 5, Appl
38	248.5	2.8	1702	4	US-08-296-791-5	Sequence 5, Appl
39	248.5	2.8	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
40	246.5	2.8	1248	2	US-08-348-353-17	Sequence 5, Appl
41	246.5	2.8	1248	2	US-08-465-965-17	Sequence 17, Appl
42	246.5	2.8	1248	3	US-08-465-966-17	Sequence 17, Appl
43	234	2.6	1848	4	US-08-296-791-6	Sequence 6, Appl
44	234	2.6	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
45	233.5	2.6	1252	2	US-08-682-517-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstreser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-728-470-10



TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-719-641-10

Query Match 4.2%; Score 374.5; DB 4; Length 1529;  
Best Local Similarity 20.6%; Pred. No. 1.9e-16;  
Matches 308; Conservative 221; Mismatches 597; Indels 369; Gaps 66;

40 TTETETIGAGAEIVSGNASFTKFTNIPITDDTTPTNSNSSSSGTAVSSEDSDDTTIT 99  
DB 176 TTYSTAEPENAINUGDI-FAKGGNI---NVRRAAIRNKGKLSAD--SVSKDKSNIYL 229  
QY 100 PDPKGGAFYNASHGVLSPFMRSGTEGSLTSEIKMTGEGGAIFSQGELLFTDLTSLTIQ 159  
DB 230 SAKEGEAEI---GGVISAQNOQAKGKLMITGDKVTLTKTGAVI----- 269  
QY 160 NNLSQLSGAIF-----GGSTLSLSGITH---ATFSCNSAEVP 194  
DB 270 -DLSGREGGETYLGDERGEGKNGIOLAKKTTLEKGSTINVSCKERGGRVIAVWGDIAD 328  
QY 195 APVKKPTPEKAOTAS--ETSGSSSSGNDVSVPSSSRAEPAANLQSHFICATATPAQ 252  
DB 329 GNINAQGSIAKTGGVFESGHDLSTGDDVIDAKELWLDPPDVSIET--LTSGRNNTGE 386  
QY 253 TDTETSPHKGSGGAIYAKGDLTADSQEVLFSINKATKGGAIFAEDKVSFENITSL 312  
DB 387 NQGYTTGDTKESPKGNSISKPTLTNSTLEQL-----RRGSYVNITA- 429  
QY 313 KVOTNGAEKGGAIYAKGDLSTQSSQSLFNSNYKGGGALYVEGGINFQDLLEIRIKY 372  
DB 430 -----NNRIYVNSINLSGSLT-----HTKRDG--VKINGDITSNENGLTI-- 471  
QY 373 NKAGTPEET--KKTLPPLAKAQAAGNADAWASSPSSGSGGATTVSDGSSSGS---DSD 427  
DB 472 -KAGSVVDVHKNTLTGTFLNIVAGDSVAF-----EREGDKARNATDAQITTAQGTIVNKD 526  
QY 428 TSE-----TVPVTAAGGGL-----YTDKNLSITNITGIIETIANNKATDV-----GGG 469  
DB 527 DKQFERNVNSINGTGKGLFIANNONFTHKFDGELNIGIVTINQTTKKDKVYWNASKDS 586  
QY 470 AYVKGTLTGTCENSHRLQFLK-----NSSDKQG-----GGIYGEDNITLSNLTGKT 513  
DB 587 YNVVSLTLNTVQKFTFIKFDVSGSNSQDLRSRRSRFAGVHFNGIGGKTNFNI-GANAKA 645  
QY 514 LF--QENTAKEGGGLFIKGTDKALMTG-----ADTSNNTGLKKR 804  
DB 646 LFKLKNATDPKKELPIT-FNANITATGNSDSSVDFDIHANLTSRAAGINMDSINITGG 704  
QY 550 -----TSEKHGGGAFVKE-----ISQYTS--DVETIPGTPVHGGETVITGN 590  
DB 705 LDPSITSHNRNSNAPEIKKDLTINATGNSFNFLKQTKDSFVNEYSKHAINSHNLITLGGN 764  
QY 591 KSTGG--NGGCVCTKRLALSNLQSLISISGNSAENGGAHTCPDSPFTADTAQEPAAASAA 649  
DB 765 VTLLGGENSSSITGNITNANKAVTLQ-----ADTSNNTGLKKR 804  
QY 650 TSTPKAPV--SPALSTPSSSTVSSLTLLAASSQASPAATSKETQDPNADTDLIDYVVD 707  
DB 805 TLTGLNISVEGNLSLGTGANAVIGNUSIAEDSTFKGEASDNL-----NITGTTNNGTAN 859  
QY 708 TTISKNTAKGGGIYAKKAMSRIDOLNISENSA-----TEIGGICCK-----ESIELD 757  
DB 860 INIKQGVVQLQDINNKG-----LNITNASGTQKTLINGNITNEKGDNLNIIKAKD 912  
QY 758 ALVSLSVTLNVKEGG-GLHAKTVNIN--LKSGFSPNNKANSSTGVTATASAPAA 813

DB 913 A--EIQIGNISQEGNLTISSDKVNITNQITIKAGVEGRSDSSAEANLANLTIOIKELK 970  
QY 814 AAASLQAA-----AAAAPSPATPTYSGVVGAIYGEKVTFSQCSG----- 854  
DB 971 LAGDLNLSGFENKAEITAKNGSOLTIGNASGKNADAKVFDKVKDKSIDTGHNVTLNSE 1030  
QY 855 --TCQFSGNOAIDNNP-----SQSSLNVQGGAIYAKYLSLSIGSSDAGTSYIFSQNSV--ST 906  
DB 1031 VKTSGSSNAGNDNSTGLTISAKDVTNNNVTSHKT-INI-SAAAGNVTKKEGTTINATT 1088  
QY 907 GKSOFTQCTAGGAIYSPFTVTLNCPATFSNN-----TASITATPKTSEDSGSSNSIK 957  
DB 1089 GSVEVTAQ--NGTIKGNITSQNVTVTATENLVTTENAVINATSGTWNISKTIDIKGIE 1146  
QY 958 DTIGGAIAGTATLSG---VSRFSGN---TADLGAALIGTLANANTPSATSGSONSITE 1009  
DB 1147 STSG-----NVNITASNTLKVSNITQDVTVPADAGAL--TTTAGSTISATTGNAN--- 1196  
QY 1010 KITLENGSIFERNQANKGAIYSPVSISKNNITNQNTSTHDSGAIYFTKDATIESLG 1069  
DB 1197 -ITTKTGDIINGKVESSSGVTLVATGATLAVGNISGNTVTTITADSGKLTSTVGSTINGTN 1255  
QY 1070 SV-----LFTGN--NVTATO-----ASSATSGONTTANYGAA 1100  
DB 1256 SVTTSQSGDIEGTISGNTVNVNTASTGDLTIGNSAKVEAKNGAATLTAESGKLTQTGSS 1315  
QY 1101 IFGDPGTTOSSODAILTLLASSGNITFNSNLSQNGQDTPASKFCGIAGYVVLSQLQAAK 1160  
DB 1316 IYSSNQTLTAKDSSIAGNINAVTLNTTGLTTTGD---SKINATSTGLINAKDAK 1372  
QY 1161 -----GKTIISPFDCVHTS-----TKKTGSTONQVYETLIDINKEENSNPYTGIVFISEL 1208  
DB 1373 LDGAASGDRTV--VNATNASGSGNVTAKTSSSVNI--TGDNL-----TINGINII 1418  
QY 1209 HENKSYIPQNAILLHNGTLVLKTELHVYVFEQEGSKLIMEPGAVLSN---QNIAH-GA 1264  
DB 1419 SEN-----GRNTVLRGKEIDVKYIQGVASVEEVIEAKRLEKVKDLSDEERETLALGV 1474  
QY 1265 LAI-----NGLTIDLSSMGTPQAGEIFSPPELRIVATTSSASGGSGVSSSIPTN 1313  
DB 1475 SAVRFEVPPNAAIVNTQN-----EFTTKPSSQVITSEKACFCSSGNGARVCTN 1522

RESULT 3  
US-08-038-682-2  
Sequence 2, Application US/08038682  
Patent No. 5549897  
GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,682  
FILING DATE: 16-MAR-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKRESSER, JERRY W  
REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-293

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-038-682-2

Query Match 4.2%; Score 373.5; DB 1; Length 1536;  
Best Local Similarity 22.2%; Pred. No. 2.2e-16;  
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TPTETIGAGAEYIVSGNASFTKFTNIPTDTPPTNSNSSSSSGETASVSDSDSTTTT 99  
Db 248 TITYSIAAPENAEVNLGDI-PAKGGNI--NVRAATIRNOCKLSAD--SVSKDKSGNIVL 301  
QY 100 PDPKGGGFYNAHSGVLSFMRSGTEGSLTSEIKWTEGGAIF-----S 144  
Db 302 SAKEGEAEI---GGVISAQNOQAKGKLMITGDKVTILKTGAVIDLSGKEGGETYLGDE 357  
QY 145 QGE-----LLFTDLTSL---TIGNLSQLSGAIFGGSTISLSGITKATFSCNSAEVAP 196  
Db 358 REGNGIQLAKKTSLEKGSTINVGKRGGRRAIVMGDIALIDGNAQGGDIAGTKGF 417  
QY 197 VKKPT-----PKAQTASSETSGSSSSGND-----SVSPSPSSR 230  
Db 418 VETSGHDLFIKDAIVDAKELWLLDFDNVINSNAETAGRSNTSEDDDEYTCGNSASPTKRNK 477  
QY 231 APAAANL-----QSHFICATAPAAQTDTEFS-----TPSHKPGSGGAIYAKGDLT 277  
Db 478 EKTTLTNTTLESILKKGTFVNITNQRIYVNSSINLSNGSLTLWSEGRSGGVEINNDIT 537  
QY 278 IADQEVLFISINKATKDGAIKPAEDVSE-----NITSLKVTQNGAEKGGAIYAKGDL 332  
Db 538 TGDDTR---GAMLTYSGWVDVHKNISLGAQGNINITA---KQDIAFEKGSQNVITQGG 591  
QY 333 STQSSKQSLFN-SNYSKGGGALYVEGGINFODLEIRIKVYKAGTFE-----TKKITLP 386  
Db 592 TITSGNQGFENNVNLTGSGS-----GLQF--TTRKTKYAITNKFEGTLNISCKVNIS 644  
QY 387 SLKAQASG-----NADAWASSPQSGGATTVSDG-----SSGSDSDTSETVPVTA 436  
Db 645 MVLKPNESGYDKFKGRTYWNLTG-----LNVSSEGFNLTIDSRGSDSAGTLTQPNL 697  
QY 437 KGGGLYTK-----NLSITNITGIEIAN-----NKATDVGGGAYVKGTLTCEN 480  
Db 698 NGISFNKDTTFNVERNARVNFIDKAPIGINKYSSLNYSFNGNISVGGSGVDFTLLASS 757  
QY 481 SH-----RLOFLKNSDKQGGIYGEONITLS-----NLT----- 510  
Db 758 SNVQPGVIVNSKYFNVSTGSSSLREKTSKGTGFSI--EKDITLNATGNNITLLQVEGT 815  
QY 511 ---CKTLFOENTAKEGGGLFIKTDKALT-MTGLDFECLINNTSEKHGGAFVTKELS 565  
Db 816 DGMICKGIVAKKNITFEGGNIIF-GSRKAVTEIEG--NVTINNANVTLLIGSDFDNHQP 872  
QY 566 QYTSDDVETIPGTPVHGETVITGKNSGGGGGVCTKRLALSLNQISISGNSAAE--- 622  
Db 873 LTIKDV-----INSGNLTAGN-----IVNIAGNLTVESNA 905  
QY 623 -----NGGGAHTCPDSEPTADTAQPAASAAATSTPKSAPVSTALSTPSSSTVSS 672  
Db 906 NFKAITNFTFNVGGL-----FDNKGNSNISIAKGAERFKDIDNSKLSITNNTSSSTYR- 958  
QY 673 LTLAASSQASPATSNKETODPNADTDLIDYVDTTTSKNTAKKGGIYAKKAKMSRID 732  
Db 959 -----TISGNTITKNG----- 970

QY 733 OLNII-SENSATE--IGGGICCKESLELDALVSLSVTENL---VGREGGLHAKTVNISNL 786  
Db 971 DLNITNEGSDTEMQIGGDVSOKEGNLTISSDKINTIKQITIKAGVDGENDSDATNNANL 1030  
QY 787 K-----SGFSFSNNKA-----NSSSTGVATTASAPAAAAASLQAAAAAPS 827  
Db 1031 TIKTKELKLTQDLNISGF---NKAETAKDGSDLTIGNTSADCTNAKV----- 1077  
QY 828 SPATPTYSGVVGGAII--GEKVTFSQCSTCOFSQNOAIDNNPSSOSSLNVQGGAIYAKTS 885  
Db 1078 -----TFNOVKDSKISADGHKVTL---HSKVETSGS---NNNTEDSSDNNAGLTIDAKN- 1125  
QY 886 LSISSDAGTSYIFSGNSVSTGK-----SOTTGOIAGGAIYSPVTLNCPPATFSNNTAST- 940  
Db 1126 -----VTYNNNITSHKAVSISATSGEI-----TTKTCTTINATTTGNVE 1163  
QY 941 ATPKTS-----EDGSSGNSIKDTIGGAIAGTAITLSGVSRFSGNTADLGAATIGTLANANT 996  
Db 1164 ITAQTGSLGGIESSGSVTLTATEGALA-----VSNISGNT-----VTVTANSQA 1209  
QY 997 PSATSGGNSITEKITLENGSFIFERNQANKGAIYSVSISIKGNNTFNQNTSTHDGSA 1056  
Db 1210 LTTLAGSTIKGTESVTTSSQS-----GDIGGTISGGTVEYKA-----TESLTTQSNK 1257  
QY 1057 IVFTRKDATIESLGSVLEFTGN-NVTATQAS--SATSGQNTN-TANYGAAIFGDPGTTQSSQ 1112  
Db 1258 IKAT-----TGEANVTSATGTIGGTIGTGNVTNVTANAGDLTVGN--GAELNA 1302  
QY 1113 TDAILTLASSGNITFNSNLSQNNQDTPASKFCGIAGYVKLSLQAAGKTTISFDFCVHT 1172  
Db 1303 TEGAATLTSSGKLT-----TEASSHITSAGQVNLQAQ--DGSVAGSINAANV 1349  
QY 1173 STKKTGTONVYETLDINKEENSNPYGTITVPSSELHENKSYIPONAILHNGTLVLKEKT 1232  
Db 1350 TLNFTG-----TLTVTKGSINATSGTLVINA-----KDAELNGAALGNHT----- 1390  
QY 1233 ELHVVSFEQKESKLIMPEPAVL-----SNQNIANGALAINGLTI 1272  
Db 1391 ---VYNATNANGS-----GSVIATTSRVNITGDLITINGLNI 1425

## RESULT 4

US-08-302-832-2  
; Sequence 2, Application US/08302832  
; Patent No. 5603938  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,832  
; FILING DATE: 16-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US pct/us93/02166

; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.2%; Score 373.5; DB 1; Length 1536;
Best Local Similarity 22.2%; Pred. No. 2.2e-16;
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TETETICEAGAEYIVGNSAFTFTNPTDTPNTNSNSSSSGETASVSDSDSTTTT 99
DB 248 TITISAAPENEAIVNLGDI-FAGGNI---NVRAAITRNQKLSAD--SVSKDKSGNIVL 301
QY 100 PDPKGGAGFYNAHSGVLSFMSTRSGTEGSLTSLBKMTGEGGAIF-----S 144
DB 302 SAKGEAEI---GGVISAQNAQKAGKLMITGDKVTLKTGAVIDLSGEGGETYLGDE 357
QY 145 QGE-----LLETDTLSL---TIONNISQLSGAIFGSGTSLSGITKATPSCNSAEVPAP 196
DB 358 RRGKNGIQLAKKTSLEKGSTINVSKEKGRAIVMGDIALIDGNINAQSGDIAKTGGF 417
QY 197 VKKPT-----PKAOTASETSSSSSGND-----SVSSPSSSR 230
DB 418 VETSGHDLFTKONAIYDAKWLDFDNVSNIAETAGRSNTSEDEYTGSGNSASTPKRNN 477
QY 231 AEPAAANL-----OSHFICATATAPAAQPTDTS-----TPSHKPGSGGAIYAKGDLT 277
DB 478 EKTTLNTTLESILKKTGFVNITANQRIYVNSSINLSGSLTLWSEGRSGGVEINNDIT 537
QY 278 IADQEVLFESINKATDGGGAIFAERKDVFE-----NITSLKVQTNGAERKGGAIYAKGDL 332
DB 538 TGGDTR---GANLTIYSGGVWVHKNISLGAQGNINITA---KQDIAFEKGSNOQVITQG 591
QY 333 SIQSSKQSLFN-SNYSKQGGALYVEGGINFQDLEERIKYKAGTPE-----TKKITLP 386
DB 592 TITSGNQKGRFNVNLSUNGTS-----GLQF--TTKRTNKYATINPKFGLNISGKNVIS 644
QY 387 SLKAQASAG-----NADAWASSPQSGGATTVSDSGD-----SSSGSDSDTSETVPVTA 436
DB 645 MVLPKNESGYDKFKGRTYNLT-----LVNSESSEFNLTIDSRGSDSAGTLTPYNL 697
QY 437 KGGGLYTDK-----NLSTNTITGIEIAN-----NKATDVGCGAYVKGTLTCEN 480
DB 698 NGISFNKQDTFFNVERNARVNFIDKAPIGINKYSLNAYSPFNISVSGGSGVDFTLIASS 757
QY 481 SH-----RLQELKNSSDKOGGIIYGEDNLTLS-----NLN----- 510
DB 758 SNVQTPGVVINSKYFNVSSTGSSURFTSGSTKTFSTI--EKDLTLNATGNTITLLQVEGT 815
QY 511 -----GKTLFOENTAKEGGGLFKIGTDKALT-MTGLDSDFLCINNTSEKHGGGAFVTKEIS 565
DB 816 DGMIGKGIYAKKNITTEGGNITF-GSRKAVTEIEG--NVINNANVTLLIGSDFDNHQKP 872
QY 566 QTVTSVETIPGTPVHGETVITGKNSGTGGNGGVCVKRLALNSLISISGNSAAF--- 622
DB 873 LTIKQDV-----IINSGLNTAGN-----IVNIAGNLTVESNA 905
QY 623 -----NGGGAHTCDPSFTADAEQPAASAATSTPKSAPYSTALSTSPSSSTVSS 672
DB 906 NFKAITNFTFNVGGL-----FONKGSNISIYAKGGAREFKDIDNSKNLITTSNSTR- 958

QY 673 LTLAASQASPATSNKETQDPNADTDLIDYVVVDTTISKNTAKKGGIYAKKAKMSRID 732
DB 959 -----TIISGNIITNKG----- 970
QY 733 QLNI-SENSATE--IGGGICCKESLELDALVLSVLTENL---VGKEGGGLHAKVTNLSNL 786
DB 971 DLNITNEGSDTEMOIGDVSQKEGNLTISSDKINIKQITIKAGVDGSDNSDATNNANL 1030
QY 787 K-----SGFSFSNNKA-----NSSSTGVATTASAPAAAAAASLQAAAAAPS 827
DB 1031 TIKTKELKLTQDLNLSGF---NKAETAKDGSDLTIGNTNSADGTNAKKV----- 1077
QY 828 SPATPYTSGVVGGAIV--GEKVTFPQSGTCQFSGNQAIDNNPSSQSLNVQGGAIYAKTS 885
DB 1078 -----IFNQVKDSKISADGHKVL---HSKVETSGS---NNNTEDSSDNNAGLTIDAKN- 1125
QY 886 LSISSDAGTSYIFSGNSVSTGK-----SQTGQIAGGAIYSPVTNCPATFSSNNTASI- 940
DB 1126 -----VTVNNNITSHKAVSISATSGEI-----TTKTGTTINATTGNVE 1163
QY 941 ATPKTS-----EDGSSGSIKDTIGGAIAGTAITLSCVSRFSGNTADLGAAGTFLAYANT 996
DB 1164 ITAQTGILGGIESSGSLTLTATEGALA-----VSNISGNT-----VTVTANSGA 1209
QY 997 PSATSGSONSITEKITLENGSFIFERNQANKRGAIYSPSVSIKGNITFNQNTSTHIDGSA 1056
DB 1210 LTTLAGSTIKTESVITSSOS-----GDIGGTISGTVVEVKA---TESLTTQJNSK 1257
QY 1057 IYTKDKATISLSGLVLTGN-NVTATQAS--SATSGQNTN-TANYGAAIFGDPGTTVSSQ 1112
DB 1258 IKAT-----TGEANVTSATGTIGGTISGTVNVTANAGDLTVGN--GABINA 1302
QY 1113 TDAILTLASSGNTTFSSNLSLONNGDTPASKFCSIAGYVKLSLOAAKGTISFFDCVHT 1172
DB 1303 TEGAATLTTSSGKLT-----TEASHITSAKQVNLQAQ---DGSVAGSINIANV 1349
QY 1173 STKTKSTQNVYETLDINKENSNPYTGTIVFSELHENKSYIPQNAIHLNGTLVLKERT 1232
DB 1350 TLNTTG-----TLTVKGSNINATSGTLVINA---KDAELNGAALGNHT----- 1390
QY 1233 ELHVVSFEQEGKGLINEPGAVL-----SNQNIANGALAINGLTI 1272
DB 1391 ---VVNATNANGS-----GSVIATTSRVNITGDLITINGLNI 1425

RESULT 5
US-08-530-198-2
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: JWB-1186  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-530-198-2

Query Match 4.2%; Score 373.5; DB 2; Length 1536;  
Best Local Similarity 22.2%; Pred. No. 2.2e-16;  
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TTEIIGEGAGAYIVSGNASFTKFNIPITDTTPTNSNSSSSGSETASVSDSDSTTTT 99  
DB 248 TITYSIAAPENAVNLGDI-FAKGGNI---NVRAATIRNOGLSAD--SVSKDKSGNIVL 301  
QY 100 PDPKGGAFYNAHSGVLSFMTSGTEGSLTSEIKMTGEGGAIF-----S 144  
DB 302 SAKEGEAEI---GGVISQNOQAQGGKLMITGDKVTLTGAVIDLSCKEGGETYLGDE 357  
QY 145 QGE---LFLTDLTSL-----TIONNLSQSGAIFGGSTISLGTATKATFSCNSAEVPAP 196  
DB 358 RGEKGNGIQLAKKTSLEKGSTINVSKEKGRRAIVWGDIALIDGNINAGGSGDIATGCF 417  
QY 197 VKKPE-----PKQATSETSGSSSSGND-----SVSPSSSR 230  
DB 418 VETSGHDLFIKDNAIVDAKEWLLDFDNVNSINAEATAGRSNTSEDDYTGSGNSASTPKRKN 477  
QY 231 ASPAAANL-----QSHFICATATPAAGTDPTS-----TPSHKPGSGGAIYAKGDLT 277  
DB 478 EKTTLTNTLESILKKGTFVNITANQRIYVNSINLSGSLTLWSEGRSGGVEINNDIT 537  
QY 278 IADSOEVLFSINKAKDGAFAEKDVSE-----NITSLKVOTNGAEKGGAIYAKGDL 332  
DB 538 TGDTR---GANLTIYSGWVDVHKNISLGAOCGNITA---KQDIAFEKGSNOVITGOG 591  
QY 333 STOSSKQSLFN-SNYSKQGGALYVEGGINFODLEIRIKYNKAGTFE-----TKKIILP 386  
DB 592 TITSGNQKGFNNVSLNGTGS-----GLQF--TTKRTNKYAITNKKFECTLNISGVNIS 644  
QY 387 SLKAQASAG-----NADAWASSQSGGATTVDSGD-----SSSGSDSDTSETVPVTA 436  
DB 645 MVLPMNESGYDKFGTYNLTG-----LNVSSEGENLTIDSGSDSAGTLTQPYNL 697  
QY 437 KGGLYTDK-----NLSTITNITGIEIAN-----NKATDVGGGAYVKGTLTCEN 480  
DB 698 NGISENKDOTTENVERNARVNFIDKAPIGIKYSSLYNASFNGNISVSGGSDVDFTLASS 757  
QY 481 SH-----RLQFLKNSDKQGGIYGEDNITLS-----NLT----- 510  
DB 758 SNVQTPGVVINSKYFNVTGSSLRKFTSGSTKTFPSI---EKDLTLNATGNTLTLQVGT 815  
QY 511 ----GKTLFQENTAKEGGGLIKGTDRALT--MTGLDSFCLINNTSEKHGGGAFVTKETS 565  
DB 816 DGMICKGIVAKKNITFEGGNITF-GSRKAVTEIEG--NVTINNANVTILIGSDFDNHQKP 872  
QY 566 QYVTSDETIPGIVTGVHGTIVTGNKSTGGNGGVCTKRLALSNLQSTISGNSAAE--- 622  
DB 873 LTIKKDV-----IINSGLNLTAGN-----IVNIAGNLAVESNA 905  
QY 623 -----NGGGAHTCPDSEPTADTAEQPAAASAAATPKSPAPVSTALSTPSSSTVSS 672  
DB 906 NPKATNTFTFNVGGL-----FDNKGNSNISIAKGAARKFDIDNSKNUSITNTSSSTVR- 958

QY 673 LTLAASSQSPATSNKETQDPNADTDLILDYVVDTTISKNKTAKKGGIYAKKAKMSRID 732  
DB 959 -----TILSGNLTNKG----- 970  
QY 733 QLN1-SENSATE--TGGGICCKESLELDALVSLSVTEMU---VGKEGGGLHAKTVNINSL 786  
DB 971 DLNITNEGSDTEMQIGGDVSOKEGNLTITSSDKINIKQTIKAGVDGENSDSDATNNANL 1030  
QY 787 K-----SGFSFSNNKA-----NSSSTGVATTASAPAAAAASLQAAAAAPS 827  
DB 1031 TIKTKELKLTQDLNLSGF-----NKAEITAKGSDJLTIGNTNSADGNTAKKV----- 1077  
QY 828 SPATPTYGVVGGAIIY--GEKVFSCQSGTCOFSGNOAIDNNPSSQSLNVQGGAIYAKTS 885  
DB 1078 -----TFNOVKDSKISADGHKVTL---HSKVETSGS---NNNTEDSSDNNAGLTIDAKN- 1125  
QY 886 LSIGSDAGTSYIFSGNSVSTGK-----SQTGQIAGGAIYSPVTVLNCPATFSNNVATSI- 940  
DB 1126 -----VTVNNNITSHKAVSISATSGEI-----TTKTGTTINATTGNVE 1163  
QY 941 ATPKTSS-----EDGSSGNSIKDTIGGATAGTAITLSGVSRFGSGNTADLGAAGITLANANT 996  
DB 1164 ITAQTGSLIGLIESSGSVTLTATEGALA-----VSNISGNT-----VTVTANSGA 1209  
QY 997 PSATGSGNSITEKITLENGSFIFERNOANKRGAIYSPSVSIKGNITFNQNTSTHDGSA 1056  
DB 1210 LTTLAGSIKGTESVTTSSQS-----GDIGGTISGGTVEVKA-----TESLTTQSNK 1257  
QY 1057 IYFTKDATIESLGSVLFNGN-NVTAQAS--SATSGQNTN-TANYGAAIFGDPGPTQSSQ 1112  
DB 1258 IKAT-----TCEANVTSATGTIGGTISGNTVNTANAGDLTVGN--GAEINA 1302  
QY 1113 TDAILTLASSGNITFNSNLSLONNOGDPASKFPCSIAGYVKLSLQAAKGTISFDCVHT 1172  
DB 1303 TEGAATLTSSKLT-----TEASSHITSAGQGVNLSAQ--DGSVAGSINAANV 1349  
QY 1173 STFKTGTQNVYETLIDINKEENSNPYGTGTIVFSSSELHENKSYIPQNAIHLNGTLVLKKE 1232  
DB 1350 TLNWTG-----TLTTVKGSINATSGTLVINA-----KDAELNGAALGNHT----- 1390  
QY 1233 ELHVVSFQKEGSKUIMEPGAVL-----SNQNTANGALAINGLTI 1272  
DB 1391 ---VYNATNANGS-----GSVIATTSRRVNTIGDLITINGLNI 1425

RESULT 6  
US-08-469-880-2  
Sequence 2, Application US/08469880  
Patent No. 5876733  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,880  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-880-2

Query Match 4.2%; Score 373.5; DB 2; Length 1536;  
Best Local Similarity 22.2%; Pred. No. 2.2e-16;  
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TPTETIGAGAEYIVGNSAFKFTNIPITDTTPTNSNSSSSGSETASVSDSDSTTTT 99  
DB 248 TTYSTAEPENEAUNLGDII-FAKGGNI---NVRATIRNOGLSAD--SVSKDKSGNIVL 301  
QY 100 POPKGGGAFYAHSGVLSFPMRTSGTSLSEIKMTGEGGAIFF-----S 144  
DB 302 SAKAGEAEI---GGVISAQNOQAKGKLMITGDKVTLTGAVIDLSKREGGETYLGDE 357  
QY 145 QGB-----LFTDLTSL-----TTONNLSQLSGAIFGCGSTISLIGTKATFSCNSAEVPAP 196  
DB 358 RGEKNGIQLAKKTSLEKGSTINVSKEKGRGIVMGDIALIDGNINAQSGGDIKTGGF 417  
QY 197 VKKPTB---GANLTIYSGGWDVHKHNSISLGAOGNINITA---KQDIAFEKGSNOQVITGOG 591  
DB 418 VETSGHDLFIKDAIYDAKEWLLDFDNVINAETAGRSNTSEDEYTGSGNSASTPRNK 477  
QY 231 AEPAAANL-----OSHFICATAPAAQDTETS-----TPSKRPGSGGAIYAKGDLT 277  
DB 478 EKTTLTNTTLESILKGTFTVNITANQRIYVNSSINLSGLTLWSESGSGGVEINNDIT 537  
QY 278 IADSQEVLFSINKATKDGGAIFAEKDVSE-----NITSLKVQTNAGAEKGGAIYAKGDL 332  
DB 538 TGDGTR---GANLTIYSGGWDVHKHNSISLGAOGNINITA---KQDIAFEKGSNOQVITGOG 591  
QY 333 SIQSSKOSLFN-SNYSKOGGALYVGGINFQDLERIKYKNAKAGTFF-----TKKITLP 386  
DB 592 TITSGNQKGRFNVSLSNTGTS-----GLQF---TTARTNKYAITNKFETLNISGRVNS 644  
QY 387 SLKAQASAG-----NADAWASSPQSGGATTVSDSGD-----SSSGSDSDTSETVPVTA 436  
DB 645 MVLPKNESGYDKFKGTYNWLS-----LNVSESGEFLNLTIDSRGSDSAGTLTQPYNL 697  
QY 437 KGGGLYTDK-----NLSTNTTIGIEIAN-----NKATDVGGGAYVKGTLTCEN 480  
DB 698 NGISFNKDDTTFNVERNARVNFIDKIPIGINKYSSLNYSFNGNISVSGGSDFTLLASS 757  
QY 481 SH-----RLOFLKNSDKOGGGIYEDNITLS-----NLT----- 510  
DB 758 SNVQTPGVVINSKYFNVSSTLSRFTSGTGTGFSI---EKDLTLNATCGNITLLQVEGT 815  
QY 511 ----GKTLFOENTAKEEGGLFKTGDKALT-WTGLDLSFCLINNTSEKHGGGAFVTKETS 565  
DB 816 DGMICKGIYAKKNITPEGNITF-GSRKAVTEIEG---NVTINNANVTLLIGSDFDHNQRP 872  
QY 566 QTYTSDVETIPGITPVHGETVITGNKSTGGNGGVCTKRLALNSLOSISISGNSAAE--- 622

DB 873 LTIKKDV-----IISGNLTAGN-----IVNIAGNLTVESNA 905  
QY 623 -----NGGAHTCPDSPPTADTAQPAASAATSTPKSAPYSTALSTPSSSFVSS 672  
DB 906 NFKAITNFTFVGGGL-----FDNKNSNISIAKGGARFKDIDNSRNLSTTTSSSYR- 958  
QY 673 LTLAASSOASPATSNKETQDPNADTDLIDYVVDVTTISKNTAKKGGIYAKKAKMSRID 732  
DB 959 -----FIISGNITKNG----- 970  
QY 733 QLNI-SENSATE---ICGGICCKRESLELDALVLSVTENL---VGKGGGLHAKTVNLSNL 786  
DB 971 DLNITNEGSDTEMQIGGDVDSQKEGNTISSDKINITKQITIRAGVDGSDSDATNNAL 1030  
QY 787 K-----SGFSFNKA-----NNSSTGVATTASAPAAAAAASLQAAAAAPS 827  
DB 1031 TIKTKELKLTQDLNISGF-----NKAEITAKDGLDITIGNTSADGTNAKV----- 1077  
QY 828 SPATPTYSGVVGAIY--GEKVTFSCSGTCQFSGNOAIDNNPSQSLNVQGGAIYAKTS 885  
DB 1078 -----TFNQVKDSKISADGHKVTL---HSKVETSGS---NNNTEDSSDNNAGLTIDAKN- 1125  
QY 886 LSISSDAGTSYIFSGNSVSTGK-----SQTQOIAGGAIYSPVTTLNCPATFSNNTASI- 940  
DB 1126 -----VTVNNITSHKAVISATSGEI-----TTKTGTITNATTGNVE 1163  
QY 941 AYPKTS-----EDGSSGSIKDFIGAIAGTATLTSVSRFSGNTADLGAAGICTLANANT 996  
DB 1164 ITAQTGSIILGIESSGSVTLTATEGALA-----VSNISGNT-----VTVTANSGA 1209  
QY 997 PSATGSONSITEKITLENGSFIFERNQANKGAIYSPSVISKGNNTFNQNTSTHGS 1056  
DB 1210 LTTLAGSTIKGTESVTTSSQS-----GDIGGTISGGTVEVKA-----TESLTTQNSK 1257  
QY 1057 IFTKDATIESLGSVLFTGN-NVTATQAS--SATSQONTN-TANYGAAIFGDPGTQSSQ 1112  
DB 1258 IKAT-----TGEANVTATGTTIGTISGNTVNTANAGDLTVGN--GAEINA 1302  
QY 1113 TBAITLLASSGNITFSSNLSQNOGDTASKFCSTAGYVKLSLOAAKGTISFFDCVHT 1172  
DB 1303 TEGAATLTSSGKLT-----TEASSHTSAKQVNLQAQ--DGSVAGSINAANV 1349  
QY 1173 STKKTGSTQNYETLIDINKENSNPYGTIVFSELHENKSYIPQNAIHLNGTLVLKEKT 1232  
DB 1350 TLTNTTG-----TLTVKGSINATSGTLVINA-----KDAELNGAALGNHT----- 1390  
QY 1233 ELHVVSFEQKEGSKLIMEPAVL-----SNQNTANGALAINGLTI 1272  
DB 1391 ---VWATNANGS-----GSVIATTSRVNITGDLITINGLNI 1425

## RESULT 7

US-08-728-470-2  
; Sequence 2, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS



STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-617-697-2

Query Match  
Best Local Similarity 4.2%; Score 373.5; DB 2; Length 1536;  
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TPTETIGAGAEIVSGNASFTFTNPTPTDTPNNSSSSSGETASVEDSDSTTTT 99  
DB 248 TITYSTAAPNEAVNLGDI-FAKGNI--NVRANIRNOCKLSAD--SVSKDKSGNVL 301  
QY 100 PDPKGGAFYNAHSGVLSFMTRSGTGLSLPLSEIKMTGEGGAIF-----S 144  
DB 302 SAKGEAEI---GGVISAQOAKGKLMITGDKVTLTKGAVIDLSGKEGETYLGDE 357  
QY 145 QGE---LLFTDLTSL---TIONNLSQLSGGAIFGGSTISLSGIYKATPCSNAEVPAP 196  
DB 358 RGEKNGIQIQAATLSLEKGSTINVSKEKGGAIVWGDIALIDGNINAQSGGDIATGGF 417  
QY 197 VKKPT-----PRAQPASETSGSSSSGND-----SVSSPSSR 230  
DB 418 VETSGHDLFIKDNAIVDAKEWLLDFDNVSNATAGRSNTSEDEYTGSGNSASTPKRNK 477  
QY 231 AEPAAANL-----QSHFTCATAPAAQPDFTS-----TPSHKPGSGGAIYAKGDLT 277  
DB 478 EKTTLNTTLESILKKGTFVNITANQRIYVNSSINLSNGSLTLWSEGRSGGVEINNDIT 537  
QY 278 IADSQEVLFSINKATRDGGAIFAEDKVSFE-----NITSLKVOTNGAEKGGAIYAKGDL 332  
DB 538 TGDQTR---GANLTIYSGGWVDVHKNIISLGAQGNINITA---KODIAFERGSNOQVITQG 591  
QY 333 SIOSSKQSLFN-SNYSKQGGALYVEGGINFQDLEERIKYKAGTPE-----TKKITLP 386  
DB 592 TITSGNQKGERFNNVSLNLTGS-----GLQF--TTTRNKYATNKFEGTLNLSGRKNIS 644  
QY 387 SLKAQASAG-----NADAWASSPQSGGATTVSDGD-----SSGSDSDTSETVPVTA 436  
DB 645 MVLPKNESGVDFKPGRTYNWLT-----LNVSSEGFNLITDRGSDSAGTLTPYNL 697  
QY 437 KGGGLYTDK-----NLSTNTITGIEIAN-----NKATDVGGGAYVKGITTCEN 480

Db 698 NGISFNKDDTTFNVARNVDFDIKAPIGINKYSSLANVAFNGNISVSGGSDVDFTLASS 757  
QY 481 SH-----RLQFLKNSDKOGGIVGEDNITLS-----NLT----- 510  
Db 758 SNVQTPGVINSKYFNVSTGSSLRFKTSGTKGFSI--EKDLTLNATGNTLLQVEGT 815  
QY 511 ----GKTLFOENTAKREGGLFIKGTDKALT-MTGLDSFCLINNTSPKHGGGAFVTKEIS 565  
Db 816 DGMIGKIGIVAKKNITPECGNITF-GSRKAVTEIEG--NVTINNANVTLLIGSDFDNHQKP 872  
QY 566 QTYTSDVETIPGTPVHGETVITCNKSTGGNGGVCVKRLALSNLQISISTSGNSAAE--- 622  
Db 873 LTKKDV-----IINSGNLTAGN-----IVNTAGNITVESNA 905  
QY 623 -----NGGAHTCPDPSFTADTAEPAAASAATSTPKSAPVSTALSTPSSSVSS 672  
Db 906 NFKAITNFTNVGGL-----FDNKGNSNISIAKGGARFKDIDNSKNLSITNSSSIYR- 958  
QY 673 LTLAASSQASPATSNKETQDPNADTDLIDYVVVDYTTISKNTAKKGGIYAKKAKMRID 732  
Db 959 -----TIISGNITNKG----- 970  
QY 733 QLANI-SENSATE-IGGICCKESLELDALVLSVTENL---VKEGGGLHAKTVNISNL 786  
Db 971 DLNITNEGSDTEMOIGDVSQKEGNLTISSDKINITKOITIKAGVDGENSDSDATNANL 1030  
QY 787 K-----SGFSFSNNKA-----NSSSTGVATTASAPAAAAASLQAAAAAAPS 827  
Db 1031 TIKTKELKLTQDLNISGF---NKAETAKDGSDLTGTNTNSADGTNAKKV----- 1077  
QY 828 SPATPYSGVVGGAIV--GEKVTFQSCGFCQFSGNOAIDNNPSSSLANVOGGGAIYAKTS 885  
Db 1078 ----TFNQVKDSKISADGHKVTL---HSKVETSGS---NNNTEDSDNNAGLTIDAKN- 1125  
QY 886 LSTGSSDAGTSYIFSGNSVSTGK-----SQTGQIAGGAIYSPVTLNCPATFSNNTIASI- 940  
Db 1126 -----VTVNNNITSHRAVSIATSGEI-----TTKTGTINATTGNVE 1163  
QY 941 ATPKTSS-----EDGSSGNSIKDTIGGAIAGTATFLSGVSRFSGNTADLGAAGTANANT 996  
Db 1164 ITAQTGILGIGESSGSVLTATEGALA-----VSNISGNT-----VVTANSGA 1209  
QY 997 PSATSGSONSITEKITLENGSIFERQNAOKRGAIVSPYSIKGNNTTFNQMTSTHGS 1056  
Db 1210 LTLTAGSTIKGTESVTTSSQS-----GDIGGTISGTTVEVKA-----TESLTTQNSK 1257  
QY 1057 IYFTKDATIESLSGVLEFGN-NVTATQAS--SATSGQNTN-TANYGAAITGDPDPTQSSQ 1112  
Db 1258 IKAT-----TGEANVTISATGTIGGTISGNTVNTANAGDLTVGN--GAEINA 1302  
QY 1113 TDAILTLASSGNITFNNNSLQNNQDTPASKFCISAGYVKLSLQAAKGTISFFDCVHT 1172  
Db 1303 TEGATLTTSSGKLT-----TEASSHITSAGOVNLSAQ--DGSVAGSINAAV 1349  
QY 1173 STKKTGTQNVYETLDINKEENSNPYTGTVFSELHENKSYIPQNAIHLNGTLVLKEKT 1232  
Db 1350 TLNNTG-----TLTVKGSINATSGTLVINA-----KDAELNGAALGHT----- 1390  
QY 1233 ELHWVSFEQKEGSKLMEPGAVL-----SNONIANGALAINGLTI 1272  
Db 1391 ---VVNATNANGS-----GSVIATTSRRVNTGDLTITINGLNI 1425

RESULT 9  
US-08-719-641-2  
; Sequence 2, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

COMPUTER: IBM PC compatible  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE:

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625

TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:

US-08-719-641-2

Query Match 4.2%; Score 373.5; DB 4; Length 1536;  
Best Local Similarity 22.2%; Pred. No. 2.2e-16;  
Matches 316; Conservative 156; Mismatches 515; Indels 437; Gaps 67;

QY 40 TTETIGEGAGAEYIVSGNASFKFNIPITDTPTNSNSSSSGSETASVSDSDSTTTT 99  
DB 248 TITYSTAAPENEAIVLGD1-FAKGGNI---NVRATIRNQGLSAD--SVSKDKSGNIVL 301  
QY 100 PDPKGGGAFYNAHSGVLSPTMSGTGLTLSEIKWTGGGAIF-----S 144  
DB 302 SAKEGAEI---GGVISAQNOQAGGKLMITGDKVTLTGAVIDLSGEGGETYLGDE 357  
QY 145 QGE---LFLTDLTSL-----TIONNLSQLSGGAIFGGSTISLTKATFFSCNSAEVPAP 196  
DB 358 RGEKNGIQLAKKTSLEKSGTINVSKEKGRAIIVMGDIALIDGNIQAQSGDIAKTGCF 417  
QY 197 VKKPT-----PKAQTASSETSGSSSSGND-----SVSPSSSR 230  
DB 418 VETSGHDLFIKNAIVDAKEWLLDFDNVNSINAEATGRSNTSEDDYTGSGNSASTPKRKN 477  
QY 231 APFAAANL-----QSHFICATATPAAGTDTEIS-----TPSHKPGSGGAIYAKGDLT 277  
DB 478 EKTTLTNTTLESILKKGTFTVNQRIIVVNSINLSGSLTLWSEGRSGGVEINNDIT 537  
QY 278 IADSOEVLFSINKATKDGGAIFAEDKVSFE-----NITSLKYOTNGAEKGAIIYAKGDL 332  
DB 538 TGDTR---GANLTIYSGGWDVHKNISLSGAQGNINITA---KQDIAPEKGSNOVITGO 591  
QY 333 STOSSQSOLFV---SNYSKQGGALYEGGINFODLEIRIKYNKAGTFF-----TKKITLP 386  
DB 592 TITSGNQKFRFNNVSLNLTGS-----GLQF---TTKRTNKYAITNKFECTLNSIGKVNIS 644

QY 387 SLKAQASAG-----NADAWASSSQSGGATTVSDSGD-----SSSGSDSDTSETVPVTA 436  
DB 645 MVLKNEGYDKFKRTYWNLT-----LNVSEGFENLTIDRGSADSAGTLTQPNL 697  
QY 437 KGGLYTDK-----NLSITNITGIIETAN-----NKATDVGGGAYVKGTLTCEN 480  
DB 698 NGISFNKDTTFNVERNARVNFIDKAPIGINKYSSLSYASFNNGNISVSGGSDVDFTLASS 757  
QY 481 SH-----CKTLFQENTAKEEGGLFIKGTDRALT-MTGLDSFCLINNTSEKHGGGAFYTKETS 565  
DB 816 DGMICKGIVAKKNITFEKNITF-CGRKAVTEIEG--NWTINNANVTILIGSDFDNHOKP 872  
QY 566 QYTSDEVETIPGIVPHGETVITGNKSTGGGGVCTKRLALSNLQISISISGNSAAE--- 622  
DB 873 LTIKKDV-----IINSGNLTAGN-----IVNIAGNLTVESNA 905  
QY 623 -----NGGGAHTCPDSFPTADTAEQPAAASAAATSTPKASAPYSTALSTPSSSTVSS 672  
DB 906 NFKAITNFTFNVGGL-----FDNRKNSNISIAKGGARFKDIDNSKNLSITNSSSTYR- 958  
QY 673 LTLAASSQASPATSNKETQDPNADTDLIDYVVDTTISKNTAKKGGGIYAKKAKMSRID 732  
DB 959 -----TISGNITNKG----- 970  
QY 733 QLNI-SENSATE--IGGGICCKESLELDALVLSVTENL---VGKEGGGLHAKTVNISM 786  
DB 971 DLNITNEGSDTEMOIGDVSQKEGNLTITSSDKINITKQITIKAGVDGENSDSDATNNANL 1030  
QY 787 K-----SGFSFSNNKA-----NSSSTGVATTASAPRAAASLQAAAAAPS 827  
DB 1031 TIKTKELKLTODLNISGF-----NKAEITAKGSDLTIGNTSADGTNAKKV----- 1077  
QY 828 SPATPTYSGVGGAIY--GEKVTFCQSGTCQFQSGQAIIDNNPSSQSLNVQGGAIYAKTS 885  
DB 1078 -----TFNQVKDSKISADGHKVTL---HSKVETSGS---NNNTESSDNNAGLTIDAKN- 1125  
QY 886 LSIGSSDAGTSYIFSGNSVSTGK---SQTQCIAGGAIYSPVTILNCPATFSNNTASI- 940  
DB 1126 -----VTNNNNITSHKAVSISATSGEI-----TTKTGTINATTGNVE 1163  
QY 941 ATPKTS-----EDSGSGNSIKDTIGGATAGTALTLSGVSRFSGNTADLGAATLANANT 996  
DB 1164 ITAQTGSLGGTSSSGSVTLTATEGALA-----VSNISGNT-----VTVTANSGA 1209  
QY 997 PSATSGSONSIPEKITLENLGSIFERNQANKRGAIYSPSVSIKGNNTFNQNTSTHDGSA 1056  
DB 1210 LTTLAGSTIKGTESVTTSSQS-----GDIGGTISGGTVEVKA-----TESLTTQSNK 1257  
QY 1057 IVFTKDATIESLGSLVFTGN-NVTATQAS--SATSGQNTN-TANYGAAIFGDPGTQSSQ 1112  
DB 1258 IKAT-----TGEANVTSGTIGGTIGTISGNTVNVNANAGDLTVGN--GAEINA 1302  
QY 1113 TDAITLLASSGNITFNSNLSQNNQDTPASKFCSTAGVVKLSLQAAKGTISFFDCVHT 1172  
DB 1303 TEGATLTSSSKLT-----TEASSHITSKAGQVNLQAQ--DGSVAGSINAANV 1349  
QY 1173 STPKTKGTONVYETLIDINKEENSNPYTGIVFSSSELHENKSYIPQNAIILHNGTLVLKKT 1232  
DB 1350 TLNTTG-----TLTTVKGSNINATSGTLVINA-----KDAELNCAALGNHT----- 1390  
QY 1233 ELHVVSFEQKESKLIMEPGAVL-----SNQNTANGALAINGLTI 1272  
DB 1391 ---VVNATNANGS-----GSVIATTSRVNITGDLITINGLNI 1425

RESULT 10  
US-08-617-697-10  
; Sequence 10, Application US/08617697

Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1600 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-617-697-10

Query Match 4.2%; Score 371.5; DB 2; Length 1600;  
Best Local Similarity 20.7%; Pred. No. 3.2e-16;  
Matches 310; Conservative 220; Mismatches 591; Indels 377; Gaps 67;  
QY 40 TPTETIGAEYIVSGNASFTKFTNPTDTPPTNSNSSSSSGETASVSDSDSTTTT 99  
DB 248 TITYSTAAPENEAINGDI-FAKGNI---NVRAATIRNKGKLSAD--SVSKDKSGNVL 301  
QY 100 PDPKGGAYNAHGVLSFWRTSGTSLTSLSEIKMTGEGGAIFSGOGLLFTDLTSLTIQ 159  
DB 302 SAKGEAEI---GGVISAQNOQAKGKMLITGDKVTLTKTAVI----- 341  
QY 160 NNLSQLSGGAIF-----GGSTISLSGKTK---ATFSCNSAEVP 194  
DB 342 -DUSGKEGGETYLGGERGEGKNGIOLAKKTTLEKSTINVSKEKGRAIVMGDIALID 400  
QY 195 APYKKPTEPKAQTAS--ETSGSSSSGNDVSVSPSSRAEPAANLQSHFICATATPAAQ 252  
DB 401 GNINAQGSIAKTGGFVETSGHDLISGDDVIDDAKELWLLDPPDDVSIEI--LTSGRNNTGE 458  
QY 253 TDTETTPSHKPGGGAIAKAGDLTADSOEVLFSINKATKDGGAIFAEDKVSFENITSL 312  
DB 459 NQGYTTGDTGKSPKPGKNSISKPLTNTSTLEQIL-----RRGSYVNITA- 501  
QY 313 KYVTNAAEKGGAIAKAGDLSIOSSKQSLFNSNYKGGGALYVEGGINFDLEEIRIKY 372  
DB 502 -----NNRIYVNSINLSNGLTL-----HTRDGG--VKINGDITSNENGLTI-- 543

QY 373 NKAGTFET--KKITLPSLKAQASAGNADAWASSPSGSGGATTVSDSGSSSSG---DSD 427  
DB 544 -KAGSVVDVHKNTITLCTGFLNIVAGDSVAF-----EREGDKARNATDAQITVAGQTTIWNKD 598  
QY 428 TSE-----TVPVTAAGGGL-----YTDKNLSITNITGIIETANNKATDV-----GGG 469  
DB 599 DKQFRFNNSVSLNGTGKGLKFIANNONNPFTHKFDGEINISGIVTINQTTKDKVKYWNASKDS 658  
QY 470 AYYKGTILTCENSHRLOFLK-----NSSDKQG-----GGIYGEDNITLSNLTKGT 513  
DB 659 YWNVSSLTLNTVQKFTFIKPVDSGSGNQDLRSRRSFAGVHENGIGGKTFFNI-GANAKA 717  
QY 514 LF--QENTAKEEGGLFIKGTDRKALMTG-----LDSFCLINN 549  
DB 718 LFLKPNAAATDPKKELPIT-FNANITATGNSDSSVMFDIHANLTSRAAGINMDSINITGG 776  
QY 550 -----TSEKHGGGAFVTK-----ISQYTS--DVETIPGTPVHGETVITGN 590  
DB 777 LDFSITSHNRNSNAFEIKKDLTINATGNSNFSLKQTSFYNEYSKHAINSNHLTILGNG 836  
QY 591 KSTGG-NGGGVCTKRLALSNLQISISGNSAAENGGAHTCPDSTFPTADTAEQPAAASAA 649  
DB 837 VTLCGENSSSITGNINITNKNVTLQ-----ADTSNNTGUKKR 876  
QY 650 TSTPKSAPV--STALSTPSSSTVSSLLTAAASSQASPATSNKE---TQDPNADTDLIDY 704  
DB 877 TLTGLNISVEGNLSLTGANANIVGNLSIAEDSTFKGEASDNLTITGTTNGTANINIKG 936  
QY 705 VDDTTISKNTAKKGGIYAKKMSRIDQLNISENSA---TEIGGGICCK-----ESL 754  
DB 937 VVKLGIDNN--KGG-----LNITNASGTOQTIIINGNITNEKGDNL-KNI 979  
QY 755 ELDAVLSVTENLVKGEGG-GLHAKTVNISN--LKSGFSSNNKANSSTGCVATVASA 810  
DB 980 KADA--EIQIGNISOKEGNLTISSDKVNITNQITIKAGVEGGRSDSSEANANLTIQTK 1037  
QY 811 PAAAAASLQAA-----AAAPSPSPATPTYSGVGGAIYGEKVTFSQCSG----- 854  
DB 1038 ELKLAGDLNISGFNKAEITAKNGSDLTIGNASGNDAKAKVTFDKVKDKSISTDGHNVTL 1097  
QY 855 -----TCQFSGNOAIDNNP-----SQSLNVQGGAIYAKTSLISGSSDAGCTSYIFSGNSV- 904  
DB 1098 NSEKVTSGSSNAGNDNSTGLTISAKDVTYNNVNTSHKT-INI-SAAAGNVTTKEGTIN 1155  
QY 905 -STGKSQTTGQIAGGAIYSPVTNCPATFSNN-----TASIAPTKTSBDESGSN 954  
DB 1156 ATTGSEVETAAQ--NGIKGNITSONVTVTATENLVTTENAVINATSTVIMSKTGIK 1213  
QY 955 SIKDTIGGATAGTAITLSG-----VSRFSGN-----TADLGAAGICTLANANTPSATSGSNS 1006  
DB 1214 GIESTSG-----NVNITASGNLTKVSNITGQDVTVTADAGAL--TTTAGSTISATTGNAN- 1266  
QY 1007 ITEKITLENGSFIFERNQANKRGAIYSPSVSIKGNNTFNQNTSHDGSIAIYFKDATIE 1066  
DB 1267 ---ITTKTDINGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTIN 1322  
QY 1067 SLGSV-----LFTGN--NVTATQ-----ASSATSGQNTNTANY 1057  
DB 1323 GTNSVTSSSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAATLAEKSLTQT 1382  
QY 1098 GAAIFGDPGTQSSQTDAILTLASSGNITFNSNNLQNOGDTSPASKFCSTAGYVKLSLQ 1157  
DB 1383 GSSITSSNGOTTLTAKDSSIAIGNANVNTLTGTLTTTGD---SKINATSGTLTINAK 1439  
QY 1158 AAK-----GKTSIFDCVHTS-----TKKTGSTQNVYETLDINKKEENSNPYTGTIVFS 1205  
DB 1440 DAKLDGAASDRTV--VNATNAGSGNVTAKTSSSVNI--TGDLN-----TINGL 1485  
QY 1206 SELHENKSIYPOANILHNGTFLVLEKTELHVSPFEQESKLIIMEPCAVLSN---QNZAN 1262  
DB 1486 NIISEN-----GRNVTURLRGKIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETIAK 1541  
QY 1263 -GALAI-----NGLTIDLSSMGTPOAGEIFSPPELRIVATTSASGGSGVSSSIPTH 1313

Db 1542 LGVSAREPVEPNNAITVNTQN-----EFTTKPSSQVTISEGKACFSGSGNGARVCTN 1592

## RESULT 11

US-08-617-697-9

; Sequence 9, Application US/08617697

; Patent No. 5977336

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Maltare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/617,697

; FILING DATE: 01-APR-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 05-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-557

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1599 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-617-697-9

Query Match 4.1%; Score 370.5; DB 2; Length 1599;

Best Local Similarity 19.9%; Pred. No. 3.7e-16;

Matches 317; Conservative 227; Mismatches 585; Indels 465; Gaps 69;

QY 30 FSRVETSSSTTETIGEGAGAYIVYSGNASFTKFTNPIPTDPTTPTNSNSSSSGETASV 89

Db 125 FNRVTSQISQLKILDSNGQVFLNPNGI-----TIGKDAIINNGFTAS--TLDI 174

QY 90 SEDSDSTT--TPDPKGGGAFYNAHSGVLSFMRTSGTEGSLTLSEIKMTGEGGAIFSQOE 147

Db 175 SNEIKARNFTLEQTKDRLAEIVNHLITV---GKDGSVNLI-----GKVKNEG 223

QY 148 LFTDLTSLTIONNLSQLSGGAI--FGSGTISLSGITATFSCNSAEVPAP----- 196

Db 224 I-----SVNGGSISLLAGOKITISDIINPTIYS---IAAPENEAINLGD 265

QY 197 -----VKKPTPEKQATSETGSSSSGNDGVSSPSSRAE-----PAAANLQSHFI 243

Db 266 IFPAKGNINVRATIRNKKLSADSVSKDGN-IVLSAKEGEAEIGGVISAQNOQA--- 321

QY 244 CATATPAAQTDTETTPSHKPGSGGAIYAKGDLTIADSOEVLFSINKATKDGGAIPAED 303

Db 322 -----KGGKLMITGDKVTLTKGAV---IDLSEKGEGETYLGSD 356

QY 304 -----VSPENITSLK-----VQINGAEKGGAIYAKGDLISQSSKQSLFNSYKQGG 351

Db 357 ERGEGKNGIQLAKTTLEKGSTINVSQ--KEGGRAIVMGDIALIDGININAAOKDIATKGG 415

QY 352 GALYVEGGINFODLEIRIKYNKAGTFFTKITL--PS-----LKAQASAGNADAWASS 404

Db 416 ---FVETSGHYLSIDDNAIVTKKWLDPENVTIEAPSASRVELGADRNSHAEVTKVTL 472

QY 405 POSGSGATTVSDS-----GSSSSGSD 425

Db 473 KKNNTSLATLTNTTISNLLKSAHVVNITARRKLTVNSSISIERGSHLILHSEGGQGGQV 532

QY 426 SPTSETVPVTAAGGGL-----YTKNLSITNITGIIETIANKKATDVGGAY----- 471

Db 533 IDKD-----ITSEGNLTIYSGWVDVHKNITLGSGLNI-----TTKEGDIAPEDKSGRNN 584

QY 472 ----VKGTILTCENSHRLOFLKNSDKQGGIYGEDNITLSNLTGKTLFOE----- 517

Db 585 LTITAGGITSONSGFR-----NNVLSNLGGKLSFTDSREDRGRRTK 629

QY 518 -NTAKEEGGLFIKGT-----DKALTMGTGLDSFCLINNTSEKHGGGAFVT 561

Db 630 GNISNKFDTGLNISGTVDISMKAPKVSWFYRDKRTYWNVT---LNVTS-----GSKF-- 680

QY 562 KEISQTYTSDVETIPGI--TPVHGETVITGNKSTGGNGGCVCTKRLALNSLOISISGNS 619

Db 681 -NLSIDSTGSGSTGPSIRNAELNG---ITFNKATFNIAQGSTANFSAIKAMPFKNANY 736

QY 620 AAEN-----GGGAHTCPDSTADTAEPAAASAATSTPKSAPVSTALSTSPSSSTVSSL 673

Db 737 ALFNEDISVSGGGS-----VNFKNASSNIQTPGVIIKSONFNVSGGSTLN-- 783

QY 674 TLLAASSQASPATSNKETQDPNADTDLIDYVVDITIS-----KNTAKGGGIY-- 722

Db 784 --LKAEGSTETAFSTIENDLNLNATGNITIROVEGTSRVNKGVAAKNITFKGNGITFG 841

QY 723 AKKAKMSRIDQLNISEN-----SATEIGGGICCKESLELDALVLSVTEHL 768

Db 842 SOKATTEIKGNVTINKNTNATLRGANFAENKSPLNAGNVINNGNLTTAGSI--INIAGLN 900

QY 769 VGKEGGGLHAKTVNISNLKSGFSFNKA-----NSSSTGVATTASAPA 812

Db 901 TVSKGANLQAITNYTFNVAG--SFDNNGASNISIRAGGAKFKPDINNTSLNITTTSDT-- 956

QY 813 AAAASLQAAAAAASPATPYSGVVGGAIFYGEK-----VTFSCSGT 855

Db 957 -----TYRTIIKGNISNKGDLNIIDKKSDAEIQIGNISKEGN 996

QY 856 COFSGNOAIDNNPSSSLNVQGG-----AIYAKTSLSIGSSD---AGTSVIFSGN---SV 904

Db 997 LTISSDKVNITNQITKAGVEGSDSSEANANITQTKELKLAGDLNLSFNKAEITA 1056

QY 905 STGKSQTTQGIAGGAIYSTVTLNCP-----ATFSNNTASIAATPKTSSRDGSGNSIKDT 959

Db 1057 KNGSDLTIGNASGGNADAKKVFDFKVKDSKISTDGHNVTLNSEVKTSGSSNAGND--NS 1114

QY 960 IGGAITAGTAITLGSVSRFSGNTADLGAATG--TLANANTPSATSGNSITBKITLENG 1017

Db 1115 TGLTISAKDVTYNN--NVTSHKTNISAAAAGNVTKEGTTINATGGS-----VEVTAQNT 1168

QY 1018 FIFERNQANKRAIYSPSVSIKG--NNITPNQNT--STHDGSAIYFTKDIATIESLGSVLT 1074

Db 1169 I-----KGNITSONVTVTATENLVTENAVINATSGVTNISTKTGDIK--GGIEST 1217

QY 1075 GNNVTAT-----QAASSATSGONTNTANYGAAIFGDPGTFQSSQTDAILTLLASSGNIF 1128

Db 1218 SGNVNLTASGNTLVKYSNITGQDVTYTDAGAL-----TTTAGST---ISATGNANIT 1268

QY 1129 SNNSLQNNQDTPAKSFCSIAGYVKLSLQAQAKGTISFFDCVHTSTKTKGTONVYETLD 1188

Db 1269 KTGDI-NGKVESSGVTLVATCATLAVGNISGNTVTITADSGKLTSTVGSTIN--GTNS 1325  
QY 1189 INKRENSNYTGTIVFESSELHENKSVIPONAILHNGTLVLKKEKTELHVVSFFQKEGS--- 1245  
Db 1326 VTTSSGSDIEGTI-----SGNTVNTASTGDLTIGNSAKV-----EAKNGAATL 1370  
QY 1246 -----KLIMEPGAVLSNQ-----IANGALA--INGLIDLSMGTPOAGEIFSPPELR 1292  
Db 1371 TAESGKLTQTGSSITSSNGQTTLTAKDSSIAGNINAANVLTNTTGT----- 1417  
QY 1293 IVATTSSAGSGSVSSIPTNPKR-----ISAAVPSGSAATPTMSENKVFEL 1339  
Db 1418 --LTTGDSKINATGCTLTINAKDKLGAAGSDRTVVNATNAGSGNVTAKTS--SSVNI 1474  
QY 1340 TGDLTLDIPNGFYQN-----PWLGSDDLVDPLIK 1368  
Db 1475 TGDLTNGLNIISENGRNTVRLRGKEIDVKYIQ 1508

RESULT 12  
US-08-728-470-9  
; Sequence 9, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berktresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-728-470-9

Query Match 4.1%; Score 366.5; DB 2; Length 1338;  
Best Local Similarity 21.0%; Pred. No. 5.2e-16;  
Matches 257; Conservative 158; Mismatches 465; Indels 345; Gaps 52;

QY 36 SSSFTTETIGBAGAEYIVSGNASFTKFTNIPPTDTTTTPTNS-----NSSSSSG 84  
Db 217 TSLTTLTNT-----TISNLLKSAHVNTARRKLTIVNSSISIERGSHLILHSEUQGG 268  
QY 85 EFASVSESDSTTTTPDPKGG-----GAFYNAH-----SGVLSFMTSGT---EBSLT 129  
Db 269 OGVIDDKDITS-----EGGNLIYSGWVDVHKHKLITLGSGLFNITTKBGLAFEDKSG 321  
QY 130 LSEIKMTGEG-----GAIFSO-----GELLTDL---TSLTIONNLSQLSGAIF 171  
Db 322 RNNLTITTAOGTTTSGNSNGFRFNNVSLNSLGGKLSFTDSREDRGRRTKGNISKNKFDG--- 378  
QY 172 GGSTISLSGITTKATFSCNSAEVPAVPVKPTEPKAQTASETSGSSSSGSDNSVSPSSSRA 231  
Db 379 ---TLNISCTVDISMKAPKVSFWYRDKRTYNNVTTLNVTSGSKF---NLSIDSTGGST 432  
QY 232 EFAAANLQSHFICATATPAAQDTTST-----PSHKPGSGGAIYA--KGDLTIAUSQE 283  
Db 433 GPSIRNAELNGI--TFNKATFNIAQGSTANFSIKASIMPFKSNANYALFNEDISVSGGS 490  
QY 284 VLFSINKATKD--GCAIFAEKDVSPENITSLKVOTNGAEKGGAIYAKGDLISQSKQS 340  
Db 491 VNFKLASSSNIQTTPGVIIKSNQFNVSGSTLNLKAEGSTETAFSI--ENDLNLAI--- 545  
QY 341 LFNSNYSKGGGALYVEGGINFQDLEIRIKYNKAGTFTETKTLPLSLKAQASAGNADAW 400  
Db 546 -----GGNITIRQVEGTSRVNK-GVAAKKNTF-----KGGNITFG 581  
QY 401 ASSPSGSGGATTVSDSGSDSDSTSETVPVTAAGGLYTDKNLSITNITGIIIEIAN 460  
Db 582 SOKATTEIKGNVTINKNTNA-----TLRGANF--AENKSPINIAGNV-INN 624  
QY 461 NKATDVGGGAYVKGTLTCENSHRLQFLKN-----SSDKQG-----GGIYGED--- 502  
Db 625 GNLTTAGSIINIAGNLTVSKGANLQAITNYTFNVAGSFDNNGASNTSIARGGAKFKDINN 684  
QY 503 ----NITL-SNLTGKTLFOENTAKEGG-GLFIKGTDKALTMFTGLDSFCLINNTSEKHGG 556  
Db 685 TSSLNITNTSDTYRTIILKGNISNKSGLDNIIDKSDAEIIOGG-----NISQEG- 735  
QY 557 GAFVTKEISQTVTSDVETIPG-ITPVHGETVITGNKSTGGNGGG-----VCTKRLLALSNO 611  
Db 736 -----NLTISSDKVNITNQITIKAG---VEGGRSDSSEAEANLTIQTKEKLKAG-- 782  
QY 612 SISISGSAE---NGGGAHTCPDSFPPTADTAEQPAAASAATSTPKSAPVSTALSTPSS 667  
Db 783 DLNISGFNKAETAKNGSDLTIGNASGGNAD-----AKKVTFDKVKDSKIST----- 829  
QY 668 STVSSLTLAASSQSPATSNKETQDPNADTDLIDYVVDVTITSKNTAKKGGGIYAKAK 727  
Db 830 -----DGHNVTLNSEVKTSGSSNAGNDNSTGLTISAKDVTVNNVNTSHKT- 875  
QY 728 MSRIDQLNISENATEIGGICCKESLELDAL-----VSLSVTE 766  
Db 876 -----INIS-----AAAGNVTTKEGTTINATTGSVEVTAQNGTIKGNITTSQNVVTATE 924  
QY 767 NLVKGEGGGLHAK--TVNIS---NLKSGFSFNKANSSTGVATASAPAAAASIAQA 820  
Db 925 NLVTTENAVINATSGTVNISTKTGDIKGGIESTSGWNVTASGNTLKV----- 973  
QY 821 AAAAAPSPATPTYSGWGGAIYERKVTFSQCSGTCQFSGNQAIIDNNPSSOSSLNVGGAI 880  
Db 974 -----NITGQDVTYTADAGALTTTAGTISATTNANITTTKGTDI 1013  
QY 881 YAKTSLSIGSDAGTSYIFSGNSVSTKSTQGTGQIAGGAIYSPVTVTLNCPATFSNNTASI 940  
Db 1014 NGKVESSGVS-----TLVATGATLAVGNISG-----NTVTI 1045  
QY 941 ATPKTSSEDSGSGNSIKDTIGGAIAGTATLSG--VSRFSGNTADLGAAG--TLANANT 996  
Db 1046 -----TADSGKLTSTVGSTINGTNSVTSSQSGDIEGTISGNTVNTASTGDLTIGNSAK 1100



Db 1149 -----NVTLNTGTLTTTGD---SKINATSGTLTINAKDAKLDGAASGD-----RTV 1192

QY 1115 AILTLASSGNITFSNNSLQNOGD 1139

Db 1193 VNATNAGSGNVTAKTSSSVNITGD 1217

RESULT 14

US-09-377-155-33

; Sequence 33, Application US/093771155

; Patent No. 6197312

GENERAL INFORMATION:

; APPLICANT: PEAK, Ian Richard Anselm

; APPLICANT: JENNINGS, Michael Paul

; APPLICANT: MOXON, E. Richard

; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

; FILE REFERENCE: 065064/0128

; CURRENT APPLICATION NUMBER: US/09/377,155

; CURRENT FILING DATE: 1999-08-19

; PRIOR APPLICATION NUMBER: PCT/AU98/01031

; PRIOR FILING DATE: 1998-12-14

; PRIOR APPLICATION NUMBER: GB 9726398.2

; PRIOR FILING DATE: 1997-12-12

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 33

; LENGTH: 2353

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-377-155-33

Query Match 4.0%; Score 361.5; DB 4; Length 2353;

Best Local Similarity 19.2%; Pred. No. 2.6e-15;

Matches 344; Conservative 265; Mismatches 684; Indels 499; Gaps 80;

QY 7 TAVFAVL-PSVSGFCFPEPEKLN-----FSRVETSSSTFTTET----- 44

Db 36 TAVLATLLFATQANATDEDELDPVVRAPVLSFHSDEKTEGEKTEVNSNNGIYFDNK 95

QY 45 -IGEGAEIVSGN-----ASFTKFTNPTTD-TTITNSNSSSSSGETAS 88

Db 96 GVLKAGAILKAGDNLIKONTDESTNASSFTYSLKDLDTLSVATEKLUSFGANGDKVD 155

QY 89 VSDSDSTTTTPDKGGGAFYNHSGVLSFMTRSG--TEGSLTSLSEIKMTGEGGAIFSG 146

Db 156 ITSDANGLAKTNGNVHLNGLDLTPDAVTNTGVLSSSFTPNDEKTI----- 205

QY 147 ELLFTDLTSLTIONNIS---QLSGGAIFGGSTISL-----SGITKATFSCNAAEVPAPV 197

Db 206 -----RAATVKDLNAGWNKIGAKTAGGNVESVDLVSAYNNVEFTTGDKNLTDVVLTA 258

QY 198 K---KPTD---PKAQTASSETS---SSSGNDS---VSSPSSSRAPAAANLOSHFICAT 246

Db 259 KENGKTEVKFTPKTSVIEKDGKLTGKENNDTNKVTNSATPDNTDEGNGLVTAKAVID 318

QY 247 ATPAAQTDFTETSPHPKSGGGAIAKAGDLTIADSOEVLFSINKATKDGGAFAEKD--- 303

Db 319 AVNKAHRVKTITANGQDFATVAGSTNVTFESGDGTASVTKDT-NGNGIIVKVDKAV 377

QY 304 ---VSPENTISLKVQNGAEKGAIAKGDLSIQSSKOSLFNS-----NYSKQGGG 352

Db 378 GDGLKFDSKKIVADTALTAVTGKVV---AEIAKEDDKKLVNAGDLVTLALGNLSWKAKA 434

QY 353 ALYVEG---GINFQDLEETRIKYNKAGTETKKTITLPSLKAQASAGNADAWASSSQSGS 409

Db 435 EADTDGALLEGIS---KQDEVKAGETVTFKAGK-----NLKVQDGGAN---FTYSLODALIT 483

QY 410 GATTVSDSGSSSGSDT---SETVPVAKGGGLYTDKNL-----SITNI- 452

Db 484 GLTSLTGLTTNGNDKAVINKDGLTITIPAGNGGTTGTNTISVTKDGKAGNKAITNVA 543

QY 453 TGI-----IEIANNKATDVG---GGAYVKGT LTC--ENSHRLQFLKNSSDRKQGGGIY 500

Db 544 SGLRAYDDANFDYLNNSATDLNHRVEDAY-KGLNLNEKNANQPLVTDSTAATVGD-- 600

QY 501 EDNITLSNLTKTLFORNTAKEEGGLFIKGTDKALTMGTGLDSFCLINNTSEKHGGAFV 560

Db 601 -----RKLGVVSTKNGTKEESNQ--VQAQDEVL-FTGAGATV---TSKSENGKHTI 647

QY 561 TKEISQT-----YTSDEVETIPGVHGETVITGNKSTGGNGGVCTKRLALSNLQSI 615

Db 648 TVSVAETKADCGLEKGDGTI---KLKVDNQNTDNVLTVCNNGTAVTK----- 691

QY 616 SGNSAAENGGGAUTCPDSDPPTAD-----TABQPAASAASTPKSAPVSTALSTPSSSTVS 671

Db 692 -----GGFETVKTGATDADRGRKVTYKDATANDADKKVATVKDVATAINS-AATFVK 741

QY 672 SLTLAASSQAAPATSNKETQDPNADTDLLIDYVVDTTISKNTAKKGG-----I 721

Db 742 TENLTTSIDEDNPTDNGKDDALKAGDT-----LTFKAGKNLKVRRDGNITFDLAKNL 794

QY 722 YAKKAKMSRIDQLNISENSATEIGGGICCKESLELDALVLSVTENLVGREGGGLHAKTV 781

Db 795 EVKTAKVS--DTLTIGNTPT-----GCTTATPKV 822

QY 782 NISNLKSGFSSNNKANSSST-----GVATTASAPAAAASLOA---AAAAAPSSUATP 832

Db 823 NITSTADGLNFAKETADASGKNVYLKGIATTLTPESSAGAKSSHVDLNVDAKKSNIA 882

QY 833 TYSVVGGAIY-----EKVTFESQS-GTCOFSGNQALDNNPSQ-----SSL 873

Db 883 EDVLRAGNIQGNNGNVVYATYDTVNFDDSTGTTVTQKADGKGVAKTGAKTISVI 942

QY 874 NVGGGAIYAKTSLSIGSS-----DAGTSYIFSGNSVS-----TGKSQT--T 912

Db 943 KDINGKLFTGCKLKDANNGATVSEDGCKDTGTGLVTAKTVIDAVNKGSGWRVTGEGAI 1002

QY 913 GQIAGGAIYSPVTLNCPATFSSNNTASIAPIKTSSEDDSGSGNSIKDTIGGAI----- 965

Db 1003 GATAVAGNAETVTSGTSVNFKNGNATTAT--VSKDNGNINVKYVDVNGDGLIGDEKKI 1060

QY 966 ---GTATILSG--VSFRSG-----NTADLGAAGICTLA-NANTPSATSGSONSIT 1008

Db 1061 VADTTTLTVTGKVSVPAGANSVNNKLVNAEGLATALANLNSWTAKADYADGESEGT 1120

QY 1009 -----EKITLENG-----SFIFE-----RNQANKRGAIYSPS 1035

Db 1121 DQEVKAGDKVTFRAGRNKLVKQSEKDEYSLQDTLTGLTITLGGTANGRNDTGV--- 1176

QY 1036 VSIKGNITFNQ-----NTSTHDGSAIYFTKDATIESLGSVLFTGN-NVTATQASSATSGQ 1090

Db 1177 INKDGTLITLANGAAAAGTASNGNTISVTKDG-----ISAGNKEITNVKSALKTYKD 1228

QY 1091 NWTAN-----YCAAI-----EGDPGTTOSQTDAILTLASSGNITFSNNSLQNN 1136

Db 1229 TQNTADTQDKEHAAVKNANEVEFVGKNGATVSAKT-----NNGKHTVTDVAEAK 1281

QY 1137 QGTPASKFCFSIAGYVKLSLQAARGKTISEFFDCVHTSTKKTGTSTQNVTYETLDINKEESN 1196

Db 1282 VDGLEK---DTDGKIKLVNDTNGNLLTVDATKGASVAKGEENAVTTDTATTAQGTNAN 1338

QY 1197 PYTGTIVFSELHENKSYIPQNAIHLNGTLVLKEKTELHVVSFEQKEGSLIMEPGAYLS 1256

Db 1339 ERGKVVVVGKS-----NGATATETDKKVVATVG---DVAKAINDAATFVK 1379

QY 1257 NONIANGALAINGLITDLDSMGTPOAGEIEFSPPELRIVATSSASGSGVSSSITPTNPKR 1316

Db 1380 VEN--DSDATIDDSPTDDGANDALKAGDGLT---LK-----AGKNLKVRRDGN 1423

QY 1317 ISAAVPSGSAATPTMSE-----NKVELTGDLTLDIPNGNFYQNPMLGSLDLDVLIK 1368

Db 1424 ITTALANDLSVKSATVSDKLSLGTNGKNKVNITSDTKCL-----NEAKDSKTGDDANIHLNG 1479

QY 1369 LPTNTSDVOVYDLTSLGSDLFPQKGYMGVTWLTDSNPQTKLOARWTFDYRRWYIPIRDNH 1428



```
QY 1036 VSIKGNITFNO-----NTSTHDSAIYFKDATIESLGSVLFTGN-NVTATQASSATSGQ 1090
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1177 INKDLGITLANGAAAGTDAENGTISVTKDG-----ISAGNKEITNVKSALKTYKD 1228
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1091 NNTAN-----YGAAI-----FGDPGTTOSSTDAILLTLLASSGNITFFSNNSLQNN 1136
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1229 TONTADETQKEFHAAVKANANEVEFKNGATVSAKD-----NNGKHTVTIDVAEK 1281
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1137 QGTPASKFCSIAGYVKLSLOAAKXTISFFDCVHTSTKKTGSTQNVYETLIDINKEENS 1196
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1282 VDGLEK---DTPCKTKLKVNDNTDGNLLTVDATKASVAKGEFNAVTTDATTAGGTNAN 1338
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1197 PYTGTVFSESELHENKSYIPQNAIHLNGTLVLKEKTELHVVSFEQKEGSKLIMEPGAVLS 1256
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1339 ERKGVVVKGS-----NGATATETDKKKVATVG---DVAKAINDAATFVK 1379
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1257 NQNIANGALAINGLTIDLSSMGTPQAGEIFSPPELRIVATTSSASGSGVSSSIPTNPKR 1316
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1380 VEN--DQSATIDDSPTDDGANDALKAGDTLT---LK-----AGKNLKVKRDGKN 1423
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1317 ISAAVPSGSAATTPTWSE-----NKVFLTGDLTLDIPNGNFYONPMLGSDLDVPLIK 1368
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1424 ITPALANDLSVKSATVSDKLSLGTNGKNVNITSDTKGL-----NFAKDSKTGDDANIHLNG 1479
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1369 LPTNTSDVQVYDITLSGDLFPQKGYMGWTWLDNSNPOTGKLQARWTFDTYRRWYIIPRDNH 1428
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1480 IASTLTD-----TLLNSGATTNLGGNGITDNEKKRAASVKD-- 1515
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1429 FYANSILGSONSMIVYKQGLINMLNN-----AREDDIAYNFWVSGVGTPLAQOQTPLSE 1484
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1516 -----VLNAGWNVYRGVKPASANNQOVENIDFVATYDTVDF-----VSG-----DKDT----- 1556
: : | | | : : | | | : : | | | : : | | | : : | | |
QY 1485 EFSYYSRGTSAIDAK---PRODFILGAAFSKIVGKTKAIKKMHNHYFHKGSE 1533
: : | | | : : | | | : : | | | : : | | | : : | | |
Db 1557 -----TSVTVESKDNKRTVEVKIGA-----KTSVIKDHNGKLFTEGKE 1593
: : | | | : : | | | : : | | | : : | | | : : | | |
```

Search completed: October 25, 2001, 08:41:01  
Job time: 350 sec

**THIS PAGE BLANK (USPTO)**



Query Match		5.1%; Score 452; DB 1; Length 2249;
Best Local Similarity		20.1%; Pred. No. 2.2e-13;
Matches		445; Conservative 274; Mismatches 822; Indels 678; Gaps 99;
Qy	29	NFSRVTSSTFTTETGEAGAEEIVSGNASFTKFTNIPPTDTPPNSSSSSSGETAS 88
Db	214	NTNALATVNVGAGTATLGGAVIK-----ATTTLTN--AASVLTLTN-----AV 257
Qy	89	VSEDSSTTTTDPKGGGAFYNAHSGFTSRSGTSLTSEIKMTGEGGAIFSGEL 148
Db	258	LTGAIDNT-----GGD-----NVGLNL-----NGALS-----QVTGDIGN----- 289
Qy	149	LFTDLTSLTIONNLSOLS--GGAIFGSGTISLGIKATPSCNSAEVAPVKKPTEPKA 205
Db	290	-----TNSLATISVGAGTATLGGAVIKAT-TTKLTDAAAVKFTNPV-VVTGAID 337
Qy	206	QTASETSGSSSSGNDVSSPSSSRABPAAN-----LQSHFICATATPAAQTDTETSTP 260
Db	338	MTGNANNGIVFTGNTSTVGNVGNATVNVGAGLLQVGGGVVVKANTINLTDNASAVT 397
Qy	261	SHKP-----GSGGAIYAKGDLTIA--DSQEVLPFSINKATKD--GGAIFAEX 302
Db	398	FTNPVVVTGAIDNTGNANNGIVFTGNTSTVTDIGNTNALATVNVGAGTATLGGAVIKAT 457
Qy	303	DVSPENITSLKVQTNGAEEKGGAIIYAKGDLISQSSKOSLFNSNYSKOGGALYVEGGI-N 361
Db	458	TTKLTNAASVLTLTNANAVLTGAI-----DNTTGGDNVGVNLN-----GALSQVGTGIGN 508
Qy	362	QDLEERIKYKNGATETTKITLPSLKAQASAGN----ADAWASSPQSGSGATTVSDS 417
Db	509	TNSLATISV--GAGT-----ATLGGAVIKATTTKLTDAASAVKFTNPVVVTGA--IDNT 558
Qy	418	GDSSGSDSDTSEVPVTKAGGGLYTKNLSITNITGIEIANKKATDVGGAIVKGTLT 477
Db	559	GNANGI-----VFTTGNSTVTC--DIGNTSLATISVGAGTATLGGAVIKATTT 606
Qy	478	CENSHRLQFLKNSDKOGGI---YGEDNITLNLTKG-----TLFOENT 519
Db	607	KLITNAASVLTLTNANAVLTGAI-DNTTGGDNVGVNLNALSQVTDIGNTNSLATISVGA 666
Qy	520	AKESGGGLFKGTOKALT-----MTG-LDSFCLINNTSEKHGGGAPVTKEIS 565
Db	667	GTATLGGAVIKATTTKITNAVSAVKFTNPVVVTGAIDSTGNANNGIVFTTGNSTVTDIG 726
Qy	566	QTYT-----SDVETIPG-----ITPVHGETVITG--NKSTGSGGGV 600
Db	727	NTNALATVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGV 786
Qy	601	CKRLALSNLQSIISGNSAA--ENGGAHTCPDPTADTAEQPAAASAATSPKSPAPV 658
Db	787	LNLGALSQVTDIGNTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVL 846
Qy	659	STAL-STPSSSTVSLLLAASSQAASPATSKETQDPNADTDLIDYVDTTISKNTAKK 717
Db	847	TGAVDNTTGGDNVGVNLNALSQVTDIGNTNS-----LATISVGAGTATL 893
Qy	718	GGG-IYAKKAKMSRIDOLINSENSATELGGGICCKESLELDALVSLVTENLVKKEGGL 776
Db	894	GGAVIKATTTKLTNAASVLTLTNANAVLTGAI-----DNTTGGDNVGV 936
Qy	777	HAKTVNINLSKGSFNSNNKANSST-----GVATTASAPAAAAAS--LQAAAAAPSSP 829
Db	937	-----LNLGALSQVTDIGNTNSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNP 992
Qy	830	ATPIYGVVGGVIAIYGEKVTFSQCSGTQCFSGNOAIDNNPSSO-----LNVQGG 878
Db	993	V-----VVTGAIDN--TGNANNGIVFTTGNSTVGNVGNATVNVGAGLLQVGG 1043
Qy	879	AIYAKTSLSIGSDAGT-----SYIFSGNSVSTGKSQTTG----- 913
Db	1044	VVKANTINLTDNASAVTNPVVVTGAIDNTGNANNGIVFTTGNSTVGNVGNATV 1103
Qy	914	-----QIAGGAIYSPTVTL--NCPA-TFSN-----NTASTATPKTSSD 949

Db	1104	NVGAGLLQVGGVVVKANTINLTDNASAVTNPVVVTGAIDNTGNANNGIVFTTGNSTV 1163
Qy	950	GSSNSIKDTIGGAGIAGTATLTSVSRFSNGTADLGA----- 986
Db	1164	GDIGNT--NALATVNVGAGITLQAGGSLAANNIDFGARSTLEFNGLDGGGKAIPYFKG 1221
Qy	987	-----AICTLANANTPSATSGSONSITEKITLENGSIFERNQ 1024
Db	1222	AIANGNAILNVNKKLTITASHLTITGTVAEINIGAGNLTIDASVGDVITLNAQINFRAR 1281
Qy	1025	-----ANKRGA-----IYSPSV-----SIKGNNTFN-----QNTSTHDG 1054
Db	1282	DSVLVLSNLTGVGVNILLAAADLVAPGADEGTVVFGVGNGLNVGSNVAGTFARNIGDGG 1341
Qy	1055	-----SAIYTKDATTESLSGVLTGN-----NVATQASSATSGQNTTAN 1096
Db	1342	NKFNLTIIYNAVTTDDVNLEGIQNLINKNADFTSSTAFNAGAIQINDATYTTDANNGN 1401
Qy	1097	YGAIFGDPGTQSSQTDAILTLASSGN--ITFSNNSLONNOGD-----TPASKFC 1146
Db	1402	LNI-----PAGNIQFAHADAQLVLQNSSGNDRITILGANIDPDNDEGIVILNSVTAGK 1457
Qy	1147	SIAGYVXLSLQAARKTISF---FDCVHTSTKTKTGSTQNVYETLIDINKEENSNPYTGIV 1203
Db	1458	TIAG-GKTEFGGAHKLTILPKGAGDCSTAGT--TFNTTNI--VLDITGOLELGATTANVV 1512
Qy	1204	FSSE-----LHENKSY--IPQNALIHNGTLVLKKEKTELH-- 1235
Db	1513	LFNDVOLTQTNIGGFLDFNAKMGVTLNNNVNAGAVQNTGNTNGLTLVLGASLNLR 1572
Qy	1236	---VVSFEQKEGSKLIMEPGAVLSNQ-----NIA 1261
Db	1573	VNGTAMUKVGNAGVTIAKGGVKIGEIOGTGNTLTLPAHENLTGSINKTGOALKLFNM 1632
Qy	1262	NGALAINGL-----TIDLSSMG-TPQAGEIFSPPELRIVATTSSASGGSVSS----- 1308
Db	1633	NGG-SVSGVVGTAANSVGDITTAGATSFASVNAKGTATLGGTTSFANTFTNTGAVTLAK 1691
Qy	1309	SITNPKRISAAVPSGSA-----TTPTMSENKVFLLG-----DL 1343
Db	1692	GSITSFAKNVTATSFVANSATINFNSLAFNSNITGGGTTTLTGANQVYTGTSFTDTL 1751
Qy	1344	TL-----IDPNG-----NFYQNPMLGSLDVLPLKL-- 1369
Db	1752	TLNTTFDGAAGSGGNILIKSGSTLDSGVSTLALVVVATNFDMN-NISPDKYFVISET 1810
Qy	1370	-----PTNTSDVQV-----YDLT-----LSGDLFPQKGYMGWTWL 1399
Db	1811	AGGLKPTSKENVKITINNDNRVDFTFDASTLTTLFAEDIAADVIDGDFAP-----GGPL 1864
Qy	1400	DSNPQTGKLO-----ARWTFDYRRVVIYPR---DNHFYANSILGSONSMI 1442
Db	1865	ANIPNAANIKSLELMEDAPNGSDARQAFNNFGLMTPLQEADATTHLIQDVVKPSSDTIAA 1924
Qy	1443	VVKGLINNMNLN-----NAREDDIAYNN-----FWVSGVTFLLAQOQTP 1481
Db	1925	VNNQVVASNISSNITALNARMKDVQSGNKGVPSSGDEMDAKFGAWIS---PFVGNATQK 1981
Qy	1482	LSEBFSY---SRGTSVAIDAKPRODFILGAFAFKIVGKTKAIKKMHNYFHKGSEYSYQA 1538
Db	1982	MCNSISGYKSDTTGTTIGFDGVSDDLALGLAYTR---ADTDIKLKN--KTGOKNKVES 2036
Qy	1539	SVYGGKELYFLNKHQHWALPF---LIQGVVSYGHIKHD-----TTLYPSIHERNKG 1588
Db	2037	NIYS---LYGLYN-----VPYENLFVEAIASYSNKRIRRRVATLTFTVGTGTANG 2087
Qy	1589	DWEDLGLIADLRISMDLKEPSKSDSKRITVYGEIYSIROKQFTEIDYDPRHFDCCAYR 1648
Db	2088	KYKSESYTGQLMAGVYTMYP---ENINLTPLAGLRYSIKDKYKE-----TGTYYQ 2136
Qy	1649	NLSLP-----VGAIVEGATMNCNLMYKIALAMPISYIYRNNPVCKYRVLSSNEA 1698

Db 2137 NUTVKGKNTNFTDGLGAKVSSNI-NVNEIVLTPELYAMVDYAFKN-----KVSADARLQ 2191  
QY 1699 GQVICVPTRTSARAEYSTQLYLG-----PFWTLXGNYTIDVGMVTLTQMTSCGARMIF 1752  
Db 2192 G-MTAPLPNPSFKQSTSDVGVGTAKHKMMEYRINYDTNIGSKYFAQGSVKYRVNF 2249

## RESULT 2

YDBA\_ECOLI STANDARD; PRT: 2003 AA.  
AC P33666; P76087; P76856; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DE HYPOTHEICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.  
GN YDBA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RA "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS6 (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000237; AAC74483.1; ALT\_SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT\_SEQ.  
DR EMBL; D90778; BAA15009.1; ALT\_SEQ.  
DR EMBL; D90778; BAA18860.1; ALT\_SEQ.  
DR EMBL; D90779; BAA18881.1; ALT\_SEQ.  
DR EMBL; X62680; -; NOT\_ANNOTATED\_CDS.  
DR EcoGene; EG11307; ydbA.  
KW Hypothetical protein.  
FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;  
  
Query Match 4.7%; Score 422; DB 1; Length 2003;  
Best Local Similarity 19.7%; Pred. No. 4.6e-12;  
Matches 394; Conservative 245; Mismatches 757; Indels 600; Gaps 86;  
  
QY 28 LNFSRVETSSSTFTTETIGEAGA-----EYIVSGNASFTKFTNIPD-----TTP 74  
Db 184 INLWQIDEANTVALEGVADGATKQYNHNGELVITGDNATVNNNGKTVTDGKDSIGTE 243  
QY 75 TNSNS-----SSGETASVSESDSTTTTDPKGGGAFYNAHSEVL- 116  
Db 244 INGNNGKVIQDGLDVGSGGHGIDITGDSATV--DNKGTMTVTDPSMGITQIDGKAIVN 301  
QY 117 ----SPMTRSGT-----EGSLTILS--- 131  
Db 302 NEGESTITNGCTGCTOINGDDATANNNGKTVVDGKDSTGTGTEINGNNGKVITODGDLVSGG 361  
QY 132 -EIKMTGEGGAIFSQGELLFTDLTSLTIQ-----NNLSQLSGGAIFGGSTISLSGIT 182  
Db 362 HGIDITGDSATVDNKGCTMTVTDPESIGIQVDGQAVVNEGE---SAITNGGTGTOINGD 418  
QY 183 KATFSCNSAEVPAPVKKPTPEKAQTASETSGSSSSGNDVSSPSSRAEPAANLQSHF 242  
Db 419 DATANNNG-----KTTVDGKDSGTETIAGNNGKVIQDGLDVGSG-----GHG 461  
QY 243 ICATATPAAQTDTETSTPSHKPGSGGAIYAKGDLTTADSQEVLFSINKATKGGGAIFAEK 302  
Db 462 IDITGDSATVDNKGCTMTVTDPESIG--IQIDGQATVNNEGESTITNGGT---GTOINGN 516  
QY 303 DVSFENITSLKVQTNGAEEKG-----GAIYAKGDLSSQSSKQSLFNSYKSGGGAIV 355  
Db 517 DATANN--SGKTVVDGKDSTGTGKIAGNIGVNDGSLTVTGGAHGVENIG----DNCTVN 570  
QY 356 VEGGINFQDLLEIRIKYNKAGTFETKKITLPSLKAQASAGNADAWASSPQSGGATTVS 415  
Db 571 NKGDIVVSDTSGIVLINGEG-----ATVSNTGDNVNSNEATGFSITVNS 615  
QY 416 DS-----GDSSGSD--SDTSETVPVTAAGGGLYTDKNLSITNITGIIETANNKATD 465  
Db 616 GKVSLAGSMQVGDFTGVDLNGNNSVTLAAK-----DLKVVGQKATGINSVGDANAYN 669  
QY 466 VGGGAYVKGTLTCENSHRLQFLKNSSDKQGGIYGED-NITL-----SNLTGKTLF 515  
Db 670 ITGNVLVDKDKTADNAAYEFF----DPSVGIVNGSDNVTLDGKLTVVDSEVTSRQSN 725  
QY 516 QENTAKEGGGLFIKGTDKALMTGLDLSFCLINNTSEKHGGGAFVTKETISQVTSVDTI 575  
Db 726 LFDGSAEKTSGLVVIGDNGTVNMG--GLELIGEKNALADGSOVTSRLRTGYSVTS-VIVV 782  
QY 576 PGITPVH--GEVITCNKSTG-----GNGGGVC-----TRIAL 607  
Db 783 SGESSVYLNGDITISGEFFLGFAGVIRVDQKALLETGSGATLTMQDIDSEFHGHTVEI 842  
QY 608 SNLQISISG-NSAANGGGAHTCPDSTFPTADTAEP-----AAASAATSTPKAP 657  
Db 843 QNLGFAFVIGENTGINSITSLQNG---KDPAPSPVILLATNGGSATNAGTITKGVTE 899  
QY 658 VSTALSTPSSST-----VSSLTLLAASSQASPATSNKETQDPNADTDLIDY---- 704  
Db 900 QHSVFENKYSTGTSNPSIFNNDVSSITGLVAQSNSTII-----NTDSGIIDLXGKS 950  
QY 705 VVDITISKNTAKGGGIYAKKAMSRID---QLNISENSATEIGGICCKESLELDVLS 761  
Db 951 VGMALAIADSTAENOGKITLDSMWVDANDTTAMRDIASNSAIDFGTGV----GVGTDSYG 1006  
QY 762 LSVTENLVGKEGGLHAKVTNTSLKSG--PSFSNKNKANSSTGVTGTASAPAAAAA 818  
Db 1007 AGKNATAINQLG-----VITYINAGMAAYGASVTINQGTINLEKNGNVDSDLA 1061  
QY 819 QAAAAAAPSPATPTYSGVVG-----GAI--YGEKVTFSQCSGTCQFSCNQ 862

Db 1062 LVGMAYVEHGTAINDOTGVININVTGQAFYNDGTGTIVNYGTICTF-----GVQC-SGNE 1116  
Qy 863 AIDNNPSSSLNVGGAIYAKT--SLTSGSDAGTSYFSGNSYSTGK----- 908  
Db 1117 -YNNDDFTSLIYTGDDPTITRSGETVTLNKAAYTDKL-AGNVNSGTLSDQDITVSSGL 1174  
Qy 909 -SQTG-----QIAGGAT-----YSPVTNLNCPATFS--N 935  
Db 1175 LENTSGGLIINLVKDKGAVIKNAGVMTNNVDVSGGILNNAAGETAQITMAGADSSLVN 1234  
Qy 936 NTASITAPKTSSEDSGSGNSIKDTIGGAICTATLTSVRSFSGNTADL---GAAIGTLA 992  
Db 1235 NTGTI--NKIVONAGVFNNS-----GSVTGRMSAGGV--FNNQTDGAIMRGAALTGTA 1284  
Qy 993 NAN---TPSATSGSONSITEKITLEN-----CSFIFERQO-----ANKRGAIYSPSVSI 1038  
Db 1285 VANNEGTVNLGSSSEGNNTGMLEYNNNSAFNNRGEFILDNDKNAVHINOSGTLY----- 1338  
Qy 1039 KGNITFNONFTSHDGSAYI-----FTKDATIE-SLGSVLFTGN----- 1076  
Db 1339 --NTGHMISNSSHNGAVNMWGGNGRFINDTIDIVSAKSLVVSANNAGDQNAFTWNOGD 1396  
Qy 1077 -----NVTATQASSATSGQNTTANYGAAIFGDPGTTQ 1109  
Db 1397 VINFDHDSASAVKVTHSNFIQNDGIMNISCTGAVAMEGDKNAQLVNNGTINLGTAGTTD 1456  
Qy 1110 SSQTDAILTLASSGNITFSNNSLONNOGDTSPASKFCSIAGYVKLSQAAGKTLISFPDC 1169  
Db 1457 TGMIGMQIDANATDAVATENNNTINIFANDSFVSLGTGVHV-----VNNGTVVITADG 1510  
Qy 1170 VHTS--TKTGTSTQNYETLTDINKIENSNPYGTGVTFVSELSHENKSYIPONAILHNGTLVL 1228  
Db 1511 VTGSLIKQGSINW-----EGMNGNGN---SSEVHYGDTLPD-----VP 1549  
Qy 1229 KEKTELHVVSFEQEGSKLIMEPGAVL--SNQIANGALAINGLTILSSMGT----- 1279  
Db 1550 KPNT--VSTGSGDEAGGSMNLLNGYVVGTVNNGSAGKLKVNNSMNGVEINTGTGTAGTAD 1608  
Qy 1280 -----PQAGEIFSPELRIVATVTSASGSGSVS-----SSIPTN 1313  
Db 1609 TTVSFDNVVSGSLNTDADAITSTSVVWTAKGSDAGNDVMTSKNAVTDVATDASVNDI 1668  
Qy 1314 PKRISAAVP-----SGSAATT-----PTMSBNKVFELTGD----- 1342  
Db 1669 AKALDAGYTNNELTSLNVGTITAEINLSALKQVSGSQATTVFREARVLSNRFSLADAPK 1728  
Qy 1343 -----LTLIDPNGFYQNPMLGSDLDVPLIKLPTNTSDVQVYDL-----TLSGDLF 1388  
Db 1729 VGNGLAFNVAAGDPRAELGNN---TEYDMLARKTIDLSESQTSMSLEYGIARLDGD-- 1782  
Qy 1389 PQGYMGTWTLDSNPOTG-----KLQARWTFDTYRRWVYIPRDNHFVANSILGSQNSMI 1442  
Db 1783 -----GAQKAGDNGVTGGYSQFTGLKHOMSFDMGNMNV-----NNALRYDVHNL 1825  
Qy 1443 VVKOGLINMLNNAFFDII--AYNFWVSGVGTFLAOG-----TPLSE-----EFSYIS 1490  
Db 1826 DSSRSIAGNTNKTADTDVKQOYLEFRSEGAKTPEPSGLKVTPYAGVKLRHTELEGGVQE 1885  
Qy 1491 RGTSVDAIDAKPRODFILGAAPFSKIVGKTKAIKKMHNYPHKGSYSYQASV-----YCKGF 1545  
Db 1886 RNAG-----DFNLN-----MNSGETAVDSIVGLKLDYACK- 1916  
Qy 1546 LYFLLNKOHGHALPFIQGVVSYGHKHDHTTLYPSIHER--NKGDWEDLGLWDLRLISM 1603  
Db 1917 -----DGWSASATLEGGPNLSYAKSQRSTASLAGAGSQHFNVDDGQKGGGINSL-TSV 1967  
Qy 1604 DLKEPSKDSKRITVY 1619  
Db 1968 GVKYSSKESLNLDAY 1983

RESULT 3

OMP\_RICJA  
ID OMPB\_RICJA STANDARD; PRT; 1656 AA.  
AC O06653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN OMPB.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH;  
RA Uchiyama T.;  
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
RT japonica";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
CC SIMILARITY).  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AB003681; BAA20138.1; -;  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.  
FT DOMAIN 1339 1656 32 KDA BETA PEPTIDE.  
FT CHAIN 528 533 POLY-GLY.  
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 4.4%; Score 397.5; DB 1; Length 1656;  
Best Local Similarity 20.0%; Pred. No. 4.9e-11;  
Matches 376; Conservative 252; Mismatches 716; Indels 533; Gaps 91;

Qy 2 KWLSA---TAVFAAVLPVSVCFCPEPEKELNFSRVETSSSTT-----PTEIGEGAGAEYI 53  
Db 10 KLISAGLVTAATASTIVASFAGSAGAAIQQN--RTTNGVATTVGVGVGDQTV--ALANVA 65  
Qy 54 VSGNASFTKFTNIPTDTPPTNSNS-----SSSGETASVSDSDSTTTTDPKGGGAFY 109  
Db 66 VAPNAVITANAN-NGINLNTFAGSNGFLSNANNLAVTVSED-----TTL-----GFIN 114  
Qy 110 NAHSGVLFSMTRSGTEGSLTL---SEIKMTGEGGAIFSQGELLFTDLFSL---TIQNMLS 163  
Db 115 NAANNANRF-----NLTLDAKGLTLITGQG-----ITNVQSAATHNAQNIWA 156  
Qy 164 QLSGGAIFGGSTISLSGITKATFSCNSAEVPAPKKPTEPK-----AQTASETSGSSSS 218  
Db 157 KFNCGAATANN--DLISGLGTTDFGAAASTLVFDLANPTQKAPLILADNALIVNGANGTL 214  
Qy 219 G--NDSVSSPSSSSRAEPAANL---QSHFICATATPAQAOTDTETSTPSHKPGSGAIYAK 273  
Db 215 NVTNGFIQVSDKSFATYKAINIGDQGFMTNATNALNLQ-----AGGTTINFN 266  
Qy 274 GDLTADSQEVLFSINKATKD-----GGAIFAEOVSFENITSLKVO----- 315

Db 267 G---TDGTRGLVLLSKNGAATDFNVGTSGGGNL---KGIIELNTVAINGOLIANAGPANAV 321  
QY 316 ---TNGAEKGG-----GAIYAKGDLISQSSKQSLFNSNYSKQGGGALYVE 357  
Db 322 ICTNNGAGRAAGFVVDNGKAATIDGGVYAK-DWVIOQA-----NAN----- 363  
QY 358 GGINFQDLLEIRIKYNKA-----GTFFE-----TKKITLPSLKAQASAGNA 397  
Db 364 GOVNRFRHVDVIGDGTAFKTAASIVAITQNSNFGTDFGNLAAQVTPVDTMTLTGNETG 423  
QY 398 DAWASSPOSGGATVSDSGD-SSSGSDSDSEVPVTA-----KGGGLYTDKNL 447  
Db 424 DA---NNPNTAGVITFAANGPLASADANAVANNITAIASGVGVVQLSGTHT-AEL 479  
QY 448 SITNITGIETIA-----NNKA---TDVGGGAYVKGLTLCENSHRLOFLKNSDKQGGIYG 500  
Db 480 RLGNAGSVFKLADGTVINGKVNQTVLGVGLAAGAITLDSATI---TGDIGNGGGAA 535  
QY 501 EDNIYLSNLTGKTL-----FOENTAKEGGGLFTKGTDKALMTGLDSF 544  
Db 536 LOSITFLANDATKTLTGLGANIISANGGTINFQAN-----GGTIKLTSTQNNIIVD----- 585  
QY 545 CLINNTSEKHGGGAFVTEISQTSYSDVETIPGLTPVHCETVIT-----GNKSTGGNGG 598  
Db 586 CDLAITADTG-----VVDASSLTNAQTLTISGTIGITGANNTTLGQFNIGSKTTLNGG 640  
QY 599 GYCTKRLALSNLQISISONSAAENGGAHTCPDFFPTADTAEQPAAASAAATSTPKSAPV 658  
Db 641 NVAINELVIGNNSVQFAHT-----YLTITTTNAACQ 673  
QY 659 SYALSTPSSSTVSSLTLLAA-----SSQASP-ATSNKETQDPNADTDLIDYVVD---TTI 710  
Db 674 GKIIENP---VYNNNTTLAAGTNLGSAAINPLAEINFGSKGARADTVLNVGEGVNIYATNI 730  
QY 711 SKNTAKKGGIYAKKAKSRIDQLAISENSATEIGGICCKESLELDALVLSVTENLVG 770  
Db 731 ITTIDNVGSFVF-----NAGG-----KNIVS 751  
QY 771 KEGGGLHARTVNISMLKSG--PFSNNKANSSSTGVATTASAPAAAAASLQAAAAAPSS 828  
Db 752 GTVGGQGNKFTVALDNGTTVKFLGNATFNGNTI-----AANSTLQ----- 794  
QY 829 PATPTYGVVGGAIYGEKVTFSQCSQTFQFSNQAIIDNNPSSSLNVQGGAIYAKTSLSI 888  
Db 795 -----ISGNYADFTASADGTGIVEF-----VNTGPIVNLKQAVPVNALKQITV 840  
QY 889 GS-----SDAGTSYIFSGNSVSTCKSQTTGQIAGAIYSPVTILNCPATFSNNTASIT 942  
Db 841 SGPGNVVVNEIGNAGNYHGAMTDTIAFENS--LGAVLFLPS---GIPFNDAGNTIPLTI 895  
QY 943 PKTSEDGSSGNSIKDTI-----GGAIAGTAITLTSVGRFSFGNTADLGAAGITLAN 993  
Db 896 KSTVGNETAEGFSVPVSVIVSGVDVSIADQVIGDQNNIVGLG-----LGSDNGIIVN 947  
QY 994 ANTPSATGSSQNSITEKITLENGSIFERNQANKGAIYSPSVSTKGNITNQNTSTHD 1053  
Db 948 APTLVAGICTINNQGTVTLSSG-----VPNTGTVYGLGTGICASK--EKQVTFETD 998  
QY 1054 GSAIYFTRDATIESLGSLVLT---GNNVTATQASA-----TSGQNTNTAN 1096  
Db 999 -----YNNLGNIIATNTINDGVTVTTGGIAGGIAGTDFDKITGLSVNGNAN 1047  
QY 1097 --YGAIAFDPGTQSSQTDAILTLASSGNITFSNNSLQNNQGDTPASKFCSIAGYVKL 1154  
Db 1048 VRFADGIF-----SNSTSMIVTTKANNGTIVYLGNAFVGNIGDS----- 1086  
QY 1155 SLQAAKGTISFFDCVHTSKTKTGSTQNVYETLDINKENSNPYTGTVFSSSELHENKSY 1214  
Db 1087 -----DTPVASVRFTGSNNGA-----GLKGNIYSQVIDFGT---YNLGI 1122  
QY 1215 IPONAILHNGTILVLEKETEL---HVVSF---EOKEGSKLINEPCAVLSNQNIANGALAING 1269  
Db 1123 VNSNVILGGSTTAINGKIDLLNTLTTFAGGTSTWGNNTSIETTLTLANGNIGHVIA-EG 1181

QY 1270 LTIDLSMGT-----PQAGEIFSPPELRIVATTSASGGSGVSSSI-----PTNPXR- 1316  
Db 1182 AQVNATTCTTTTINVODNANANESGTQ-----TYTLIQGARENGTLGCPNFTVTSNRF 1236  
QY 1317 -----TSAAVPSGAATPTPMSENKVELTGLDGLTIDPNGNPNYONMGLSGDLVDPLIKLPIN 1372  
Db 1237 VNYGLTIRAAANQDYVITRTNNAEN--IVTNDIT-----NSPFGGAPGVQN-----VTTFVN 1285  
QY 1373 TSDVQVY-DLTLSGDLFPQGYMGWTWLTDSNPQTGLQARWTFDYYRRWYIIPRDNHFA 1431  
Db 1286 ATNTAYNNLLAKNASADSANFVGTIVTDTSAITNAQLDVAKD-----IQALGNRUGA 1340  
QY 1432 NSILGQNSMIVVYKGLINNM--NNAREDDIAYNNFWVSGVGTFLAQOCTPLSEBSFY 1489  
Db 1341 LRYLGTPEMVGSEAGAIPAVAAGDEAVDNVAY-GIMAKPEYTDHAKSKKGLAGYAK 1398  
QY 1490 SGTSTVAIDAKPRQDFILGAAFSKIVGKTKAIKKMHNYFHKHSE-----YSQASVYGGKFL 1546  
Db 1399 TTGIVIGLDTLANNLMIGAA-----IGITKTDIK-HQDYKKGDKTDVNGFSFSLYGAQ-- 1451  
QY 1547 YFLLNKHQHWALPELQI--VVSYGHKHKDTTTLVPSIH-----BRNKGDWEDLGLWADL 1599  
Db 1452 QFVEN-----FRAQSAIFSLNOVKNSQRYFPDANGNMSKQIAAGNYDNMTFEGNL 1503  
QY 1600 RISMDLKEPSKSDSKRITVYGELEY-----SSIROKQFTEIDYDPRHFDJAY 1647  
Db 1504 TVGYDY--NMQGVLTPTMAGLSYLKSSDENYKGTGVTVANKQVNS-----KFSD--- 1551  
QY 1648 RNLSLPVGCAGEGAINN 1664  
Db 1552 -RTDLIVGAKVAGGTMN 1567

RESULT 4  
Y96\_YEAST  
ID Y96\_YEAST STANDARD; PRT; 1140 AA.  
AC Q04893;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL I13.1 KDA PROTEIN IN PRES-FET4 INTERGENIC REGION.  
GN YMR317W OR YMR9924.09.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Z54141; CAA90835.1; -.  
DR SGD; S0004936; YMR317W.  
KW Hypothetical protein; Repeat.  
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 4.4%; Score 395.5; DB 1; Length 1140;

Best Local Similarity 20.6%; Pred. No. 3.7e-11;

Matches 279; Conservative 214; Mismatches 515; Indels 349; Gaps 55;

QY 36 SSSTTFTTIGEAGAEYIVSGNASFTKFTNIPTDTPPTNSNSSSSSGETASVSED--- 92



Db 127 VIFENNTCCR-----LFTWRNPYAADKIREGGA THAONLY INNHVDVVGFMKN--- 174  
QY 798 NSSSTGVATTASAPAAAAAALQAAAAAPSPPATPTYSYGVGGAIYGEKVTFSQCSGTCQ 857  
Db 175 -----FSYVQGGAI-----STANTEV 190  
QY 858 FSGNQA-----IDNNPSQSLNVGGAIYAKTSLSTGSSDAGTGYIFSGNSVSTGKSQTTG 913  
Db 191 VSENOSCFMFDMNICIQTNTAGKGAIGAYAGTNSFNSNCDLFFI-----NNA 238  
QY 914 QIAGGAIYSPVTVL---NCPATFSNNTASIAPTKTSSEGGSSGNSIKDITIGGAGIAGTAT 970  
Db 239 CCAGAGIYSPICSLTGNRGNIVFYNNRC-FKNVETASSEASOGGAIKV-----TRLD 290  
QY 971 LSGVRSFGNTADLGAAGICLANANTPSATSGSONSITEKITLENGSFIFERN-QANKRG 1029  
Db 291 VTG-----NRGRIFSDNITKNYGG 310  
QY 1030 AIYSPSVSIKGNIT-FNONTSTHDSAIYFYTKDATIESLGS-----VLFTGNNTVATQA 1083  
Db 311 AIYAPVTVLVDNGPTYFINNIANNKGAIYI--DGTSSKISADRHAIIFENIVT----- 364  
QY 1084 SSATSQNTNTANYGAAIFGDPGTTQSSQTD--ILTLASSGNITFSNNSLQNNQGDTP 1141  
Db 365 -----NVTNAN-----GTSTAMPNPRNALTIVASSSGEILLGAGSSON----- 402  
QY 1142 ASKFCSIAGYVKSLLQAARKTISFFDCVHTSTKKTGTQNVYETIDINKEENSNPYTC 1201  
Db 403 -----LIFYDPIEV-----NAGVSFNFKEADQ---TGS 429  
QY 1202 IVF-----SSELHEN--KSYIPQNAIHLNGTLVLKEKTELHVVSFEQKEGSKLIMEPGA 1253  
Db 430 VYFSGATVNSADFHORNLOTKTPAPLTLSNGFLCIEDHAQLTVNRFTQ-TGGVVSGLNGA 488  
QY 1254 VLS-----NQNTANGALAINGLTIDLSSMGTPQAGEIFSPPELRIVATTSASGSGSVS 1307  
Db 489 VUSCYKNGTGDSSASNASITLKHIGLNLSSI-LKSGAEI---PLLWVEPTNNNN----- 538  
QY 1308 SSIPNPKRISAAPVSGSAATPTMSKNVFLTDLTIDPNGNFYQNPMGLSDLDVPLI 1367  
Db 539 -----YTADTAATFSLSDVK-----LSLIDDYGN---SPYESTDLTHALS 575  
QY 1368 KLP-----TNTSDVOYVDLTLSGDLFPQKGYMGWTWLD-----SNPQRG 1406  
Db 576 SQPMLSISEASDNQSENIDFSLGNVPHYWGQLTWGWAKTQDPEPASSATITDPQKA 635  
QY 1407 KIQARWTFDYRRWYIPRDNH---FYANSILGSONSMITVVKQGLNNMLNARNAPDDIAY 1463  
Db 636 NREHRTLLLTLPAGTVPSPKRSPLIANTLWG---NMLLATESL-----KNSAELTPSGH 688  
QY 1464 NNFW---VSGVGTFLAQOQTPLSEEFYSYSRGTSVAIDAKPRQDFTLGAAFSKIYVKTKRA 1520  
Db 689 -PFWGITGGGLCMYVQDPRENHPCGFHMSSGYSGAMIAQOHTFSL--KFSQTYTKLNE 745  
QY 1521 IKKMNIYFHKGSEYSYQASVYGGKFLYLLNKHQHWALPFLQGVVSYG-HIKHDTTTLTY 1579  
Db 746 RYAKNNVSSK--NYSQ-----GEMLFSL---QEGFLLTKLV-GLYSYGDHCHHFYTCG 794  
QY 1580 PSIHNRKGDWEDLGLWDLRLSMDLKEPSKDSKRITV--YGELE-YSSIRKQKQFTETD 1636  
Db 795 ENLTSGTFRSQTMGG-----AYFFDLPMPKPGFTSHLTAPFLGALGIYSSL--SHFTYEG 848  
QY 1637 YDPRIHFD-DCAVRNLSLPGVCAVEGAIMNCLIMYNKLA---LAYMPSIYRNPNCKVR 1691  
Db 849 AYPRFSTKTPNLINLVPIG--VKGSFMNAT---HRPQANTVELAYQPVLYRQEPGIAAQ 903  
QY 1692 VLSS 1695  
Db 904 LLAS 907

RESULT 6  
OMP\_B\_RICPR

ID OMPB\_RICPR STANDARD; PRT; 1643 AA.  
AC Q53020; Q9ZCM0;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN OMPB OR SPAP OR SPA OR RP704.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsia; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=BREINL;  
RC STRAIN=BREINL;  
RC MEDLINE=91045972; PubMed=2122457;  
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
RT "Characterization of the gene encoding the protective paracrystalline-  
RT surface-layer protein of Rickettsia prowazekii: presence of a  
RT truncated identical homolog in Rickettsia typhi";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BREINL;  
RA Moron C.G., Yu X.J., Walker D.H.;  
RT "Sequence analysis of ompB of Rickettsia prowazekii";  
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MADRID E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sichert-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria";  
RL Nature 396:133-140(1998).  
RN [4]  
RP PARTIAL SEQUENCE.  
RC STRAIN=BREINL;  
RX MEDLINE=921114896; PubMed=1370573;  
RA Ching W.M., Carl M., Dasch G.A.;  
RT "Mapping of monoclonal antibody binding sites on CNR fragments of  
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
RT prowazekii";  
RL Mol. Immunol. 29:95-105(1992).  
RN [5]  
RP IDENTIFICATION OF CLEAVAGE SITE.  
RX MEDLINE=92104668; PubMed=1729180;  
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
RT membrane protein of rickettsiae: identification of an avirulent mutant  
RT deficient in processing";  
RL Infect. Immun. 60:159-165(1992).  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M37647; AAA26390.1; ALT\_INIT.  
DR EMBL; AF161079; AAD42234.1; -.

DR EMBL; AJ235273; CAA15140.1; .  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
FT VARIAT 257 257 V -> A (IN STRAIN BREINL).  
FT VARIAT 1010 1010 Y -> D (IN STRAIN BREINL).  
FT VARIAT 1450 1450 A -> S (IN STRAIN BREINL).  
FT CONFLICT 178 179 AA -> VC (IN REF. 1).  
FT CONFLICT 191 201 TTEAPLTLGA -> INSRSSSYHLVS (IN REF. 1).  
FT CONFLICT 212 212 T -> I (IN REF. 1).  
FT CONFLICT 313 313 Q -> L (IN REF. 1).  
FT CONFLICT 1104 1104 T -> G (IN REF. 2).  
FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.18; Score 371; DB 1; Length 1643;  
Best Local Similarity 21.06; Pred. No. 8.1e-10;  
Matches 373; Conservative 203; Mismatches 642; Indels 562; Gaps 87;

QY 2 KWLSTAVFAAVLPSVSGFC-FPEPKELNFRSVETSSSTFTETIG---EAGAEYIVSGN 57  
DB 10 KIISAGLVTAIIIVGFSVGAAMQYRNTNAAATTF-DGIGFDQAAGANIPVAPN 68  
QY 58 ASFTKFTNIPTDTPPTNSNS-----SSSGETASVSDSDSTTTTPDPKGGGAFYNAHSG 114  
DB 69 SVITANANNPITFNPCHLNSFLDANDLAVTINEDT----- 107  
QY 115 VLSFWTRSGTSGSLTSEIKMTGGGAIFS---OGELLFTDLTSLTTONNLSQLSGGAIF 171  
DB 108 TLGFTITIAQAQKF----FNFVTAAGKILNITGGGITVQEAASNTINAQNALTKVHGGAII 163  
QY 172 GGTISLSGITKATFSCNSAEVPAPVKPTEPKAQTASETSGSSS---SSGNDVSVPSS 228  
DB 164 NAN--DLSGLGSIIFAAAPSVLLENLNP---TQEPAPLTLGANSKIVNGNGTL----- 213  
QY 229 SRAEPAANLQSHFICATAPAAQTDTETSTPSHKPGSGGAIYAKGDLTIADSOEVLFSI 288  
DB 214 -----NITNGFIQVSDNTFAGIKT-----INIDDCOGLMFN- 244  
QY 289 NKATKDGGAIFAEDKVSFENTISLKVQTNGAEKGGAIYAKGDLISQSSKOSLFSNSYSK 348  
DB 245 --STPDAA-----NTLNQVGGNTINFNGIDGTGKLVLVSKNGAATFENVTG 289  
QY 349 QGGGALYVEGGINFODLEEIRIKYKAGTFETKITTPLSKAKAAGNADAWASSSPQS 408  
DB 290 TLGNL---KGI-----IELNTAA-----VAGKLISQGAANAVTGTONGAGRA 330  
QY 409 SGATTVSDSGSDSDSDTSETVPVTAKGGGLYTDKNLSITNITGIEIANNKATDVGG 468  
DB 331 AGFIVSDVNGNAATIS-----QQVYA-KNMVIOQANAGGQVTFEHIVDVGL 375  
QY 469 GAYVKGILTICENSRLQFLKNSDKGGGIIGEDN-----ITLSNLTKTL 514  
DB 376 G-----GTFNFKTADSKVIITENS-----FGSTNFGNLDFTQIVVPDPTKILKGNFIGDVK 425  
QY 515 FOENTAK-----EEGGGLFIKGTDKALMTGLDSDFLCLINNTSEKHGGGAFVTKESQVTS 570  
DB 426 NNGWTAGVITFNANGALYSASTDNIAVTNI-----NATEAEGAGVV---ELSGIHIA 475  
QY 571 DVEITPG--ITPVHGETVITG--NKSTGGNGGGVCTKRLALNSLQSIISGNSAAENG 626  
DB 476 ELRLNGGGSIFKLADGTVINGPVNQNALMNNALAAAGSIQLDG--SAIITGDI---GNGG 530  
QY 627 AHTCPDSEPTADTAEOAPAAASATSTPKSAPVSTALSTPSSSTVSSLTLLAASSQSPAT 686  
DB 531 VNAALQHITTLANDASKILALDGANII--GANVGGAIHFQANGGTIKET----- 576  
QY 687 SNKETQDPNADTDLIDYVVDITISK-----NTAKKGGGIYAKKAKMSRIDOL 734  
DB 577 -----NTQNNIVNFDLIDITDKTGVDVDASSLTNQTUTINGSIGTVVANTKTIAQL 628  
QY 735 NISENSATEIGGGICCKESLELDALVLSVTLENLVGREGGLHAKTVNISN----- 785

DB 629 NIG--SSKTILNAGDVAINELVIENNGSVQLNHNIT-----LITKTINAANQOIIIVAAD 681  
QY 786 -----LKSG-----FSPSNKANSSS-----TGV---ATTASAPAAAASL 818  
DB 682 PLANTNTTLAGDTNLGSAENPLSTHFEATKAANADSIILNVGKGVNLVANNITNDANVGS 741  
QY 819 QAAAAAAPSPATPYTSGVVGGAIVGEKVTFSQCSTGTCFQSGNQAIIDNPSSOSL---NV 875  
DB 742 HF-----RSGGTSIVSGTVGGQ--QGHKL-----NNLILDNGTIVKFLGDTTF 782  
QY 876 QGGA-IYAKTSLSGSDAGTYSYIFSGNSYSTGKSOTTGQIAGGAIYPTVTLN----- 928  
DB 783 NGGTKEGKSILQI-SNNYTTDHVESADNTGTLEFVNTDPI-----TVTLKQAGAYF 833  
QY 929 -----CPATFSNN-----TASIAPTKTSSEGGSGNS-----TKDT 959  
DB 834 GVLQKQVITSGPNIVFNEIGNVGIVHGAANASISFENASLGTSFLPSGTPDLVLTIKST 893  
QY 960 IGGAIAGTA-----ITLGSVSRFSGNTADLGAAGITLANANTPSATSGSONSITEKI- 1011  
DB 894 VGN--GTVDNFNAPIVVVSGIDSMINN---GQIIGD--KKNIIALSGLSDNSITVNAN 944  
QY 1012 TLENGSFIERNQA-----NKGAIYSPSVSIKGNITFNQNTSTHDGSAIYFTKD 1062  
DB 945 TLYSGIRTKNNOGTVTLSSGMPNPGTIY--GLGLENGSPKLKQVFTTDD----- 993  
QY 1063 ATIESLGSVLTGNNVTATQASSATSGONTNTANYGAAI-----FGDPGTQTS 1110  
DB 994 --YNNLGSII--ANNVTINDYVTLTGTGIAGT-DFDAKITLGSYNGNANVRFPD--STFS 1046  
QY 1111 SOTDAIITLLASSGNITFSNNSLONNOG--DTPASKFCSTAGYVKLSLOAAKGKTISFFD 1168  
DB 1047 DPRSMIVATQANKTVTYLGNALVSNIGSLDTPVA----- 1081  
QY 1169 CVHFTSKTKGTQ-----NVY-ETLDINKEE-----NSNPYT--GTIVFSSELHENKSYI 1215  
DB 1082 ---SVRFTGNSGAGLQGNISQNIQDFTGYNLNLNSNVILGGTGAINGEI-----DLL 1133  
QY 1216 PQNALHNGTILVLEKTE---LHVVSFEQKEGSKLMEPGAV-----LSNQNIAN 1262  
DB 1134 TNNLIFANGTSTWGDNTSISTTLNVSS--GNIGQVIAEDAOVNATTTGTTTIKIQDNAN 1191  
QY 1263 -----GALAINGLTIDLSSMGTPQ---AGEIFSPPEL----- 1291  
DB 1192 ANFSCTQATYLIQGGARENG-----TLGAPNFAVTSNIFVKYELIRDSNQDYVLFTN 1245  
QY 1292 RIVATTSASGGVSSSIPNPKRIISAAVPSGSAATPTTMSKNKVFLTGDLTLDIPNGN 1351  
DB 1246 DVLNVVTTAVGNSAIAA-PGVSONISRCLES-----TNTAAYNNMLIAKD----- 1290  
QY 1352 FYQNPMLGSLDVLPLIKLPTNTS---DVQVYDLTSLCDLFPQKGYMGCTWLDNSNPQTK 1407  
DB 1291 -----PSDVATFVGAIAATDTSAAVTTVNLNDTKQDQLLSNR--LGTLYLSNAETSD 1341  
QY 1408 LQARWTFEDTYRRWYIIPRDNHFYANSILG---SONSMIVVKOGLINMLNNAFDDIAYNN 1465  
DB 1342 V-----AGSATGAVSSGDEAEVSYGVWAKPYNYNLAEOQDK--- 1376  
QY 1466 FWVSGVGTFLAQOQTPLSEEFYSYRGTSTVAIDAKPRODFILGAAFSKIVGKTKAIKKMH 1525  
DB 1377 ---GGIAGYKAK-----TTGVVVGDLTSLASDNLMIAGAA---IGITKTDIK-H 1416  
QY 1526 NYFHKGSEYSQA---SVYGGKFL-----YFLLNK 1552  
DB 1417 QDYKKGDKTDINGLSFLSYGSQLVKNFFAQCNAIFTLNK 1456

RESULT 7  
OMP\_RICRI  
ID OMPBL\_RICRI STANDARD; PRT; 1654 AA.  
AC Q53047;  
DT 30-MAY-2000 (Rel. 39, Created)





Db 840 HPMFDRIALGAACNAGVTHNMSFY-GTKKSKGKGFHSTTLGASLR--CELKRDMSPLRS 896

QY 1614 KRITVGELEYSSIROKQFTEIDYDRPHFD-DCAYRNLSLPVGCAGEAIGMNCNLMYN- 1671

Db 897 IMLTPPAQLFSRTEPASIRESCDLARLFLEQAHAVVSPIG--IKGA-----YSS 946

QY 1672 -----KLALYMSPIYRNPNVCKYVLSSNEAGVGVPTRTSA-----RAEYSTQLY 1720

Db 947 DTWPTLSWEMELAYQPTLYWKRPLLN--TLIIQNGSNWVTTNTPLAKHSFYGRGSHSLKF- 1004

QY 1721 LGPFWTLYGNYTIDVGMNLTLSQMTSCGARMIF 1752

Db 1005 --SHLKLAFANYQAEVATSTVSHYNAGGALVF 1034

RESULT 9

MSB2\_YEAST STANDARD; PRT; 1306 AA.

AC F32334;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE MSB2 PROTEIN (MULTICOPY SUPPRESSION OF A BUDDING DEFECT 2).

GN MSB2 OR YGR014W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=92383951; PubMed=1514328;

RX Bender A., Pringle J.R.;

RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.";

RL Yeast 8.315-323(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=97435481; PubMed=9290212;

RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;

RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae chromosome VII.";

RT chromosome VII.";

RL Yeast 13:1077-1090(1997).

CC -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).

CC -!- SIMILARITY: SOME, TO YEAST HKR1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M77354; AAA34798.1; -.

DR EMBL; Z72799; CAA96997.1; -.

DR PIR; S25370; S25370.

DR SGD; S0003246; MSB2.

KW Transmembrane; Glycoprotein; Repeat.

FT DOMAIN 698 816 7 X 17 AA TANDEM REPEATS.

FT REPEAT 698 714 1.

FT REPEAT 715 731 2.

FT REPEAT 732 748 3.

FT REPEAT 749 765 4.

FT REPEAT 766 782 5.

FT REPEAT 783 799 6.

FT REPEAT 800 816 7.

SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;

Query Match 4.0%; Score 353.5; DB 1; Length 1306;

Best Local Similarity 19.0%; Pred. No. 3.9e-09;

Matches 257; Conservative 237; Mismatches 516; Indels 343; Gaps 48;

QY 153 LTSLTIONNLSQSG-CAIFGGSTISLSGITKATFSCNSAEVPAPVKKPTPEKQAASET 211

Db 9 LSTLVISGLSARSPDFIFNGCT-----QQAQSQSES 41

QY 212 SGSSSSGNDNSVSSPSSRAEPAAANLQSH----FICATATPAAOQTDTETSTPSHKPGSG 267

Db 42 QGOVSETNEASQDSSTTSLVTAYSQGVHSHQSATIVSATISSLPSTWYDASSTQTSVS- 100

QY 268 GAIYAKGDLTIADSOEVLFSINKATDGGAIKFAEKOVSPENITSLKVOT--NGAEKGGGA 325

Db 101 ---YA-----SQESDYAVNONWSA-----STNQLPSTSTSYAYAPTSTSDAFRASS 145

QY 326 IYAKGDLS-----IQSSKQSLFNSNYSKQGGALYVEGGINFODLEIRIKYNKAGTFT 380

Db 146 VNAASDVSTASVPIDTSANS-----IPFTTTSNIETTSAPLISDT 186

QY 381 KKITLPSLKA-----QASAGNADAWASSPOSGGATTVSDSGSSGSD---- 425

Db 187 PLISTSTMSAADNVFSSANPISASLTITDSSFDQSTAGAIPIVQSSADFPSSSEILVQ 246

QY 426 --SDTSE---TPVPTAKGGGLYTKNLSTNITGIIETANKKATDVGGAYVKGTLTC 478

Db 247 SSADFPSPSPPTTIDISLAAPLOTSESSFTTASAAPVS---STDVGGSS---ASPVV 300

QY 479 ENSHRLQFLKNSSDKQGGIYGEDNITLSNLTKTLFOENTAKEEGGLFIKGTDKALTM 538

Db 301 SMSNAGQIASSSS-----TDNPTMSETFSLTSE-----VDGSDVISTV 339

QY 539 TGL-----DSFCLINNTSEKHGGGAFV---TKEISQTYTSDV-----ETIPGI 578

Db 340 SALLSAPFLQTSNSFSIVSPSV-----SFVPQSSSDVASSSTANVSSSFSDPQP 393

QY 579 TPVHGETVITGNKST-----CGNGGGVCTKRLSNLQSLISIGNSAAENGGA 627

Db 394 TSTSGSVSVVAQSASALAFOSSTEVYGCASASTMSSLLSTLSLQSTLTLDSSSLASSASS 453

QY 628 HTCPCDPSPTADTAEQPAAASAAATPKSPAPVSTALSTPPSSSTVSSLLTLLAASSQASATS 687

Db 454 SDLTD-YGVSTASIPLLS-----ASEQASTSSSFVSPVSPVSVFSQS 496

QY 688 NKETQDPAOTDILLIDYVVDTTTSKNKAKGGGIYAKKAKMRIDQLNISENATEIGGG 747

Db 497 SSVASTSAPS-----VSSFSFYTSLOAG-----SSMNPSSSTIVYSSSTG--- 540

QY 748 ICCKESELDALVSLVTENLVGKEGGLHAKTVNINLSKSGFSPNNKANSSTGVATT 807

Db 541 -SSEESAASATATLGGSSSTY--MAGNLQSQPPSTSLILSESOATSTSAVLASSSVSTT 597

QY 808 ASAPAAAAAALQAAAAAPSPATPTYGVVGAIFYGKVT-----YSQSTALQTSFASSTEGSETS 848

Db 598 SPYTAGGASTEASSLISSTSAETSQVS-----YSQSTALQTSFASSTEGSETS 650

QY 849 ---FSQCSTGTCQFSGNOAIDNNPSSQSLNVGGAI-----YAKTSLSIGS--SDAGTSVIFS 900

Db 651 SQGFSTSSVLVQMPSSISSEFSQTTTQMNSSASSSQYTTISSTGILSQVSDTSVSYTTS 710

QY 901 GNSVS-----TGKSQTTQGIAGGIYSPVTNCPATFSNNTASIA----TPKTSS 947

Db 711 SSVSVQVSDTPVSYTSSSSVSQVS-----DTPVSYTSSSVSQVSDTPVSYT 759

QY 948 EDGSSGNSIKDTIGGAIAGTAITLGSVRFSGNFTADLGAAGTILANANTPSATSGSQNSI 1007

Db 760 TSSSSVSQVSDT-PVSYTSSSSVSQVSDTSVPSTSSRSSVSQVSDTPVPSTSSRSSVSQ 818

QY 1008 T-----EKITLENGSFIFERNQANKGAIYSPSVISIKGNNTTFNONTSTHDSGA 1056

Db 819 TSSSLQPTTSSQRFITSTHGALSSESSVSQQAASEITSSINATASEVHSIQTTAAATQ3TT 878

QY 1057 IYTKDATIES-----LGSVLFTGNVVTATQASSATSQNTNTANYGA 1099

Db 879 LSFT-DANSSASAPLEAVATSTPTPSSKASSLLLTTPSTSLSOVATNTNVQTSLTTESTT 937





```
QY 1088 -SQONTANTYGAALFGDGTQSSQTDAILTLASSGNITFNNSL-----Q 1134
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1328 SKSSLSLSPVSSILMSQFSSSSSSSSSLASLSISPVDVTVSVLPQPTTSIATLCT 1387
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1135 NNQDTPAKFC--SIAGYKLSLQAAGKTIISFPDCVHTSTKTKGSGNQNVETLDINKE 1192
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1388 DSQCQOEVSITGNCNDDVTSTATPPSTVTDVMTCTGSECQKTTSS-----SCDGYSC 1442
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1193 ENSNYCTGTIVES-----SELHENKSYIPQNAIHLNGTLVLKKEKTELHVVSFE 1240
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1443 KVSEYTKSATISAGSGGCGQASATSEL--NSQYVTMTSVI--TPSAITTSVEVH----- 1494
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1241 OKEGSKLIMEPCAVLSNONGALAINGLITLSSMGTPOAGEIFSPPELRIVATTSSA 1300
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1495 STESISITVYKPVYTTSSDTNGEL-----ITITSSQ 1527
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1301 SGGSGVSSSIPNPKRIISAAVPSGSAATPTMSNKNVLTGDLT 1344
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1528 T----VIPSVTTIIRTKVAITSAPKPTTTTIVVEQRLSSSGIAT 1567
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 12

```
ICEN_XANCT
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ICE NUCLEATION PROTEIN.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas
OX NCBI_Taxid=343;
RN [1]
```

SEQUENCE FROM N.A.

RX STRAIN=X565;

RX MEDLINE=91080859; PubMed=2259339;

RA Zhao J., Orser C.S.;

RT "Conserved repetition in the ice nucleation gene inax from

Xanthomonas campestris pv. translucens.";

RL Mol. Gen. Genet. 223:163-166(1990).

CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE

CRYSTALLIZATION IN SUPERCOOLED WATER.

CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).

CC -!- DOMAIN: CONTAINS 133 IMPERFECT REPEATS OF THE CONSENSUS

OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A

REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.

CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE

NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN

FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; X52970; CAA37140.1; -.

DR PIR; S11672; S11672.

DR HSP; P06620; IINA.

DR InterPro; IPR000258; -.

DR Pfam; PF00818; Ice\_nucleation; 81.

DR PRINTS; PR00327; ICENUCLEATN.

DR PROSITE; PS00314; ICE\_NUCLEATION; 57.

KW Ice nucleation; Repeat; Outer membrane.

SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

Query Match 3.8%; Score 341.5; DB 1; Length 1567;  
Best Local Similarity 22.0%; Pred. No. 1.8e-08;  
Matches 276; Conservative 185; Mismatches 537; Indels 259; Gaps 56;

```
QY 67 PTTDTTTTNSNSSSSSGETASVSDSDSTTTTPDPKGGGAFYNAHSGVLSFWMTSGTEG 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 PSTQATPSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQ 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 SLTSL--EIKMTGGGALFSGGELLFTDL-----TSLTIQNNLSQSGGAFFGGSTLSL 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 S-TLGAQDSRLVAGYGTETAGD--HSDLIAGYGTGTAGSDSSILAG-----YGSTQTA 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 SGITKATFSCNSAEVPAPVPKPEKAPTASSETS-----GSSSSSGNDS--VSSPSSR 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 AGRSLTAGYGS-----TQTAQBSRLTSCYGTSTATSGSDSAVISGYGSTQ 306
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 AEPAAANLQSHFICATPAAQTDTETSTPHKPGSGGA-----IYAKGDLTITADSOEVL- 285
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 TAGSESSLTAGY--GSTQTAARKGSDITAGYGTGTAGSDSALIAGYGTGTAGSESSLT 363
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 --FSINKATKOGGAI--FAEKDVSFENITSL----KVQTNCAEEKGGAIYAKGDLSTQS 336
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 AGYGTQTAARKGSDVTAGYGTGTAGADSTLIAGYGTGTAGGESSLTAGYGTGTARQG 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 SKQSLFNSNYSKQGGGALYVEGGINFQDL-----EERIKNYKAGT 377
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 SDITAGYGTGTAGADSTLIAGYGTGTQSGDSSLTAGYGTGTQTAARKGSDITAGYGTGT 483
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 378 FETKKITLPSLKAQASAGNAD-----AMASSPQSGGATTVSDSDSSS---- 422
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 AGSDSSLIAGYGTGTQTAGSESSLTAGYGTGTQTAQDSSLTGTGYGST--STAGHDSLLIAG 541
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 -GSDSDTSETVPVTAAGGGLYT-DKNLSLTNTITGIIETANNKATDVGGAIVKGLTTCEN 480
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 YGSTGTAGYDSTLIAGYGTGTQTAQDSSLTGTGYGSTGTAGADSTLIAG---YGSTGTAGS 598
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 481 SHRL-----QFLKNSSDKQGGGIYGEDNITLSNLTQKTLFOENTAKEEGGLFIKGTD 533
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 DSSLTAGYGTGTAREGSDVTAG--YGS-----TG-TAGADSTLIAGYGTGTQSGSD 647
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 534 KALTWTGLDLSFCLINNTSEKHGGGAFVTKISQTVT--SDVETIPGTPVHGETVITGNK 591
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 648 SSLT-AGYGS-----TQTARK--GSDVTAGYGTGTAGADSTLIAG---YGSTGTQSGD 695
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 592 S--TGSNGGGVCTKR-----LALSNLOSISTSGNSAAENGGAHTCPDSPPTADT 639
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 696 SSLTAGYGTGTARKGSDVTAGYGTGTAGADSTLIAGYGTGTQSGDSSLTAGYGTGT 755
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 640 AEPAAAAAATSTPKSAPVSTAL-----STPSSSTVSSLTLLAASSQASPATSN-----K 689
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 ARKGSVDVTAGYGTGTAGADSTLIAGYGTGTQSGDSSLTAGYGTGTQTAARKGSDITAGY 815
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 690 ETQDPNADTDLIDYVDVTI-----SKWTAKKGGGIYA--KKAKMSRIDQLNI 736
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 STGTAGADSTLIAGYGTGTQSGDSSLTAGYGTGTAREGSDVTAGYGTGTAGADSTLI 875
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 737 SENSATEIGGGICCKESLELDALVSLVTENLVGEGGGLHAKTVNLSNLSKGSFSSNNK 796
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 876 SYGSTQTAGS-----DSSLTAGYGTGTQTAARKG-----SDVTAGYGTGT 915
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 797 ANSSST-----GVATTASAPAAAAAALQAAAAAPSSPATPTYSGVVGGAIFYGEKVFESQ 851
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 916 AGADSTLIAGYGTGTQSGDSSLTAGYGTGTQTAARKGSDMTAGY-GSTGTA--GADSTLIA 972
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 852 CSGTCQFSGNQADINNPSSQSLNVGGGAIYAKTISLIGSSDAGTSYIFSGNSVSTGKSGT 911
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 973 GYGSTQTSQ-----SDSSLT----AGYGTQTAAREGSDVTAGY--GSTGTAGADST 1017
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 912 TQGIAGGAIYSPTVTLNCPATFSNNNTASTATPKTSE----DGSSGNSIKDTIGGAIACT 967
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1018 L--IAG---YGSTGTAGSDSSLTAGYGTGTARQSDVTAGYGTGTAGS-----T 1065
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 968 AITLSGVSRSFGNTADLGAAIGTLANANTPS-----ATSGSQNSITEKITLNGSF 1018
```

```
Db 1066 LIAGYSTQTAGSDSLTAGYSGTQTARQSGDITAGYSGTGTAGADSSLIAGY----GST 1121
QY 1019 IPERNOANKRGALYSVSIKGNITF---NONTSTHDSGSAIYFTKDATIESLGSVLEF- 1074
Db 1122 QTAGYDSNLTAGYSGTQTAREDSLSLTAGYSGTSTAGHDSLSLIAGYSGTQTAGYNSILTTG 1181
QY 1075 -GNVTVTAQASSNTSGO-NTNTANYGAAIFGDPGTTQSSOTDAILTLASSGNITFSNNS 1132
Db 1182 YGSTQTAQESSLSLTAGYSGTSTAGYSDSLTAGYSGTQTAGYKSTLTLAGYSGNSTAGHSS 1241
QY 1133 LQNNQDTPASRFSCTAGYVKLQAKGKTIISFFDCVHTSTKTKTSTQNVYETLDI 1189
Db 1242 LIAGYST-----QIAGY-ESTLTAGYSSSL-----TQOONSSLTAGYSGSTEI 1283

RESULT 13
ID OMPB_RICTY STANDARD; PRT; 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: Identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04661; AAB48987.1; -
```

```
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 3.8%; Score 339.5; DB 1; Length 1645;
Best Local Similarity 19.0%; Pred. No. 2.3e-08;
Matches 377; Conservative 245; Mismatches 709; Indels 655; Gaps 92;

QY 2 KWLSTATVFAAVLPVSYGFC-FPEKELNFSRVETSSST-----FTETIGEAGAEIVS 55
Db 10 KIISAGLVASTATIVAGFCVGVAMGVQYNRTNAAATVVDGAGFDT--GAGVNLPA 67
QY 56 GNASFTKFTNPTTDTTPTNSNS-SSSGETASVSDSDSTTTTTPDPKGGGAFYNAHS 114
Db 68 TNSVITANSNNAITFTPNGLNSFLDNTANTLAVTINENTTL-----G 111
QY 115 VLSFMTSGTEGSLT---SEIKWTGEGGAIFSGOGLLFTDLTSLTIQNNLSQSGAIF 171
Db 112 FVTNVTQGNFFFTIGAGKSLTITGHG--ITAQ-----QAATTKSAQNVYSKVNA:AAI 164
QY 172 GGSTISLSGITTKATFSCNSAEVPAPVKKPTPEKAQTASSETSGSSSSGNDVSSPSSRA 231
Db 165 NDN--DLSGVGSIDFTAAPSVLEFNINPTTQEA----- 196
QY 232 EPAAANLQSHFICATATPAQTDTSTSPSHKPGSGAIYAKGDLTIADSOEVLFSINKA 291
Db 197 -----PLTLDGNAKIVGANGI 213
QY 292 TKDGGAIPEAKVDSPENITSLKVOTNGAEKGGAIYAKGDLTSSQSKQSLFNYSIQGG 351
Db 214 LNITNGFVKVSDRTFAGIKTINI-----GD-----NQGLMFT--TPDAA 251
QY 352 GALLYVEGGINFQDLEEIRIKYKAGTFETKKITLPSLKAQASAGNADAWASSPQSSGA 411
Db 252 NALNLOGGN-----TINFN--GRDGTGLVLVSKNGNATEFN----- 287
QY 412 TTVSDSGDSSGSDSDTSETVPVTAKGGGLYTDKNLSITNITGIE-----IANN 461
Db 288 -----VTGSLGG-----NLKGVIEDFTTAAAGKLIANG 315
QY 462 KA-----TDVGGGAVYKGLTCTENSHRLOFLKNSDKQGGGIYGEDNITLS-NLTGKTL 514
Db 316 GAANAVIGTDNGAGRAAGFTVSDVNGNAATI-----SCQVYAKDIVIQSANAGQVT 367
QY 515 FOENTAKEEGGLFIKGTDKALPTMTGLDSECLINNTSEKGGGAFVTKESQTYTSIVET 574
Db 368 FEHLVDVGLGKTNFKTADSKVIITENASF-----GSTDFGN 404
QY 575 IPGITPVHGETVITGN-----KSTGGNGGVCT-----KRLALSNLQSTISIG 617
Db 405 LAVQIVVPNNKILTGNFIGDAKNNGNTAGVITFNANGTLVSGNTDPNIVVYTNKAIIEVG 464
QY 618 NSAEANGG--GAH-----TCPDSPTAD-----TAEQPAASAATSPKSPAPVS 659
Db 465 AGIVQLSGIHGAELRLGNAGSIFKLADGTVINGPNQNLVNNNALAAGSIQDGSAILI 524
QY 660 -----TALSTPSSS--TVSSLTLAASS-----QASPATSNKETQDPNAD 697
Db 525 GDTCGNVANAALODITLANDASKILTLSGANIIGANAGAIHFQANGGT-----IQLTSTQ 580
QY 698 TDLIDYVVDVT-----ISKNTAKGGGIYAKKAKMSRIDQLNISENSATEIG 745
Db 581 NNILVDFDLDTVTQTCVVDASSLTNNQTLTNGSICITICANTKTLCGRFNVGSSKTLNA 640
QY 746 GGTCCCKE-SLELDALVSLSVTENLVKGGEGGLHAKTVNISNL----- 786
Db 746 GGTCCCKE-SLELDALVSLSVTENLVKGGEGGLHAKTVNISNL-----
```

```
Db 641 GDVAINELVMDGSGVHLTHNTYLI-----TKTINAANQKIIIVAADPINTDTALAD 692
Qy 787 -----KSGFSSNNKANSS-----TCGV---ATTASAPAAAAAASLQAAAAAPSS 828
Db 693 GTNLGASPLNIIHFATKAANGDSILHIGKGVNLYANNITTTDANVGLHP-----RS 746
Qy 829 PATPTYSGVVGAIYGEKVTFSQCSTQCFQSGNOAIDNPS---QSSLNVOGGA-IYAKT 884
Db 747 GGTISVGTGGO-QGLKL-----NNLIDNGTIVKFLGDTFNGGTKEGKS 793
Qy 885 SLISGSSDAGTSYIFSGNSVSTGKSQTTGQIAGGAIYPTVTLNCPATF-----933
Db 794 ILQI-SSNYITDHIESADNTGTFLEVNTDPI-----TVELNKQGAYFGLVKQVMVSG 844
Qy 934 -----SNNTA-SIATPKTSSDGGSGNS-----IKDTIGGAIAGTA-- 968
Db 845 PGNIAPNEIGNVGHAIADVDSISFENASLGASLFLSTGPLDLVLTIKSTVGN---GTVDN 901
Qy 969 -----ITLSGVSRFSGNTADLGAAGITLANANTPSATSGSONSIT-EKITLNGSFIFER 1022
Db 902 FNAPILVSGIDSMINN---GOVIGD-QKNIIALSGLSONSIIVNSNTLYAGIRTTKT 955
Qy 1023 NOA-----NKGAIYSPSVSTKGNNTFNQNTSTHDSGAIYFTKDATIESLGSLVF 1073
Db 956 NQGTVTLSGGIPNPNPTIY--GLGLENGDPKLKQVTFTTD-----YNNLGSIIA 1002
Qy 1074 TGNVNTATOASSATSQNTNTANYCAAIFGD-----PGTTOSSOTDAILTLASSG 1124
Db 1003 T--NTVINDVLTLTGGTAGTDFDKITLGSINGNANVKFVDRTFSHTSMIVSTKANOG 1060
Qy 1125 NITFNSLQNNQO--DTPASFCESIAGYVVKLSLQAARKTISFPDCVHTSTKTKGQSN 1182
Db 1061 TVTYLGNALVGNIGSDIPVASVRFTGNDSGVGLQGN-----IHQNIDFGT--- 1107
Qy 1183 VYETLDIINKENSNPYGTIVFSPSELHENKSYIPONAILHNGTLVLEKTELH-VVSEQ 1241
Db 1108 -YNLILASDVLG--GGTTAINGEI---DLTLNLIIFANGTSTWGNNTSLSTTLNVS 1160
Qy 1242 KEGSKLIMEPGA-----VLSNON-----IANGALATNGLTIDLSSWG 1278
Db 1161 GNVGQIVIAEGQVNAVTTGTTTIKIQNANANFSGTQTYLTIQGGAFNG-----TLG 1214
Qy 1279 TPQ-----AGEIFSPPEL-----RIVATTSASGGSGVSSSIPTNPKRISAA 1320
Db 1215 APNEDVTGNIPVKYELIRANDANQDYVLRTNDVLNVVITAVGNSAIAANA-PGVHQNIAIC 1273
Qy 1321 VPSGSAATPTMSENKVFLTGDLTLIDPNGFYQNPMLGSLDVLPLKLPNTW----SDV 1376
Db 1274 LES-----TDTAAYNNMLLAKD-----SSDVATFIGAATDTGAAVATV 1312
Qy 1377 QYVDLTLSGDLFPQKGYMGTWFLDSNPOTGKLQARWTFDTYRRWYIPRDNHFYANSILG 1436
Db 1313 NLNDQTKQDOLLGNR--LGLARYLSNSETA-----DVG 1344
Qy 1437 SONSMIVVKQGLINMLNNAAREDDIAYNFWVSGVGTFLAQOQTPLSBEFYSYRGTSVA 1496
Db 1345 SET-----GAVSS--GDEAIDQVSY-GVWAKPYNIAEQDKGLAGYKAKTAGVVVG 1394
Qy 1497 IDAKPRQDFILGAATSKIVGKTKAIKKMHNYPHKSEYSYQA---SVYGGKFLYELLANKQ 1553
Db 1395 LPTLANDNLMIGAA-----IGITKTDIK-HODYKKDKDITKGLSFLSYCAQOL----- 1442
Qy 1554 HGWALPFLIQ--VVSYGHIKHDTT-LVPSIHERNK---GDWEDLGLWDLRISMDLK 1606
Db 1443 ---VKNFAQGSATFLTNKVKSKSORYFFDANGKNKQIAGNYDNITFGNLMFGYDY- 1498
Qy 1607 EPSKDSKRITVYGELEYSSIRKQFTEIDVDPRH-----FDDCAYRNLSLPVGCAREG 1660
Db 1499 --NALQGVLVTPMAGLSYLKSSNENYKGTGTIVANKRIHSKFSQ---RIDLIIVGAKVTG 1552
Qy 1661 AIMNCH-ILMYNKLALAYMPSIYR-----NNPVCKYRVVLSNEAGQVIGCVPTTSARA 1713
Db 1553 SAMNINDIVY-----PEIHSFVHVHVNGKLSKAQSMLDGQTAPFI-SQPDRT-AKT 1602
```

```
Qy 1714 EYSTOL 1719
Db 1603 SYNIGL 1608

RESULT 14
CBPA_CLOCL STANDARD; PRT; 1848 AA.
ID CBPA_CLOCL
AC P38058;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1565642;
RX MEDLINE=92228810;
RA Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -!- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULASE
ENZYMES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC -----
CC EMBL; M73817; AAA23218.1; -.
CC PIR; A44140; A44140.
CC HSSP; Q06851; INBC.
CC InterPro; IPR001956; -.
CC InterPro; IPR002102; -.
CC Pfam; PF00942; CBD_3.1.
CC Pfam; PF00963; Cohesin; 9.
CC Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
CC SIGNAL 1 28
CC CHAIN 29 1848
CC DOMAIN 29 189
CC FT CELLULOSE BINDING PROTEIN A.
CC FT CELLULOSE-BINDING (BY SIMILARITY).
CC FT SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;
CC SQ
```

Query Match 3.7%; Score 334.5; DB 1; Length 1848;  
Best Local Similarity 20.4%; Pred. No. 4.7e-08;  
Matches 380; Conservative 265; Mismatches 771; Indels 443; Gaps 80;

```
Qy 4 LSATAVFAVLPSVSGFCFPEPKELNF-----SRVETSSSTTFTETIGAGAIYVSGNAS 59
Db 11 LALMVFALVLPISIPALAATSSMSVEFYNSNKSAGTNSITPIKITNTSDSL----NLN 66
Qy 60 FTKETNIPTDITTTPT-----NSNSSSSGGETASVSESDSTTTTPDPKGGGA 107
Db 67 DVKRYRYTSDGTQCTQCTWCDHAGALLGNYSYDNTSKVTANFVKETASPTSYD----- 120
Qy 108 FYNAHSGVLSFMTRSGTEGLTSLSEIKMTGEGGAIFSGCELLTDLTSLTIQNLLS---- 163
Db 121 -----TYVEFGFASGRATLK-----KGQFITQGRITKSDWSNVTQNDYSFSDAS 165
Qy 164 -----QLSGGAIFG---GSTI---SLSGITKATFSCNSAEVPAPVK----- 198
Db 166 SSTPVVNPVKTYIGGAKVGLGTAPGPDVPSIINPTSATFDKNVTK-QADVKTMTNLGN 224
```



[4]  
RN CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.  
RP TISSUE=Submaxillary gland;  
RA MEDLIN=97248516; PubMed=9092502;  
RX Gerken T.A., Owens C.L., Pasumarthy M.;  
RT "Determination of the site-specific O-glycosylation pattern of the  
FT porcine submaxillary mucin tandem repeat glycopeptide. Model proposed  
FT for the polypeptide:galnac transferase peptide binding site.";  
RL J. Biol. Chem. 272:9709-9719(1997).  
CC -!- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN  
CC SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY  
CC GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE  
CC THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL  
CC ENVIRONMENT.  
CC -!- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A  
CC MULTIMERIC MUCIN STRUCTURE.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.  
CC -!- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81  
CC RESIDUES.  
CC -!- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF  
CC THE REPEAT UNITS, HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER  
CC RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE  
CC GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE  
CC PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO  
CC ENHANCE GLYCOSYLATION.  
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M61883; AAA30998.1; -;  
DR EMBL; M21174; AAA30990.1; -;  
DR PIR; A40009; A40009;  
DR InterPro; IPR000359; -;  
DR InterPro; IPR001007; -;  
DR Pfam; PF00007; Cys\_knot; 1.  
DR PROSITE; PS01185; CTCK\_1; 1.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS01208; VWFC; 1.  
KW Glycoprotein; Repeat. 1  
FT NON\_TER 1  
FT DOMAIN <1 368 81 AA TANDEM REPEATS.  
FT REPEAT <1 44 1.  
FT REPEAT 45 125 2.  
FT REPEAT 126 206 3.  
FT REPEAT 207 287 4.  
FT REPEAT 288 368 5.  
FT REPEAT 369 391 6 (INCOMPLETE).  
FT DOMAIN 929 995 VWFC.  
FT DOMAIN 1062 1146 CTCK.  
FT DISULFID 1062 1109 BY SIMILARITY.  
FT DISULFID 1076 1123 BY SIMILARITY.  
FT DISULFID 1085 1139 BY SIMILARITY.  
FT DISULFID 1089 1141 BY SIMILARITY.  
FT DISULFID ? 1145 BY SIMILARITY.  
FT CARBOHYD 46 46 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 50 50 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 51 51 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 57 57 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 58 58 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 61 61 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 66 66 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 67 67 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 73 73 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 74 74 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT CARBOHYD 76 76 O-LINKED (GALNAC. .) (MUCIN TYPE).  
FT \*

FT	CARBOHYD	77	77	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	81	81	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	83	83	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	87	87	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	91	91	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	93	93	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	94	94	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	96	96	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	98	98	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	101	101	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	103	103	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	104	104	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	106	106	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	107	107	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	108	108	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	110	110	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	114	114	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	117	117	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	123	123	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	124	124	O-LINKED (GALNAC. .)	(MUCIN TYPE).
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	547	547	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	917	917	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	985	985	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1002	1002	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1068	1068	N-LINKED (GLCNAC. .)	(POTENTIAL).
SQ	SEQUENCE	1150 AA;	109615 MW;	3CB68B5D29DD7F5A	CRC64;

Query Match 3.6%; Score 325.5; DB 1; Length 1150;  
Best Local Similarity 20.2%; Pred. No. 6.5e-08;  
Matches 267; Conservative 148; Mismatches 451; Indels 453; Gaps 57;

Qy	70	DTTPTNSNSSSSSETASVSDSDSTTTTPDKGGGAFYNAHSGLVFWFRSGTEGSLT	129
Db	1	ETARP-----SVAGSGTTGTVSGAGSGTSGSSSGSTG-----ATGASIGQPE	41
Qy	130	LSEIKMTGEGGAIFSGQELLFTDLTSLTIQNNLSLGGAIFG-GSTISLSGIKATFSC	188
Db	42	TSRTSVAGSSGA-----PAVSSGASQAAGTSGAGPGTTASSVGVTEA---84	
Qy	189	NSAEVPAPVKKPTPEKQAOTASSETSGSSSSGNDSSVSSPSSRAEPAANLOSHFICATAT	248
Db	85	-----RPSVAGSGTTGTVSGAGSGTSSS-GSPG-----AT114	
Qy	249	PAAQDTETSTPSHPKPGSGGAIYAKGLTIADSQEVLFSINKATKDGGAIFAEDVSPEN	308
Db	115	GASIGQPETSRIS-VAGSSGA-----134	
Qy	309	ITSLKVQTNGAEEKGGAIFYAKGDLISQSSKQSLFNSNYSKOGGGALYVEGGINFQDLBEI	368
Db	135	-----PAVSSGASQAAGT-----SGAGPGT-----154	
Qy	369	RIKYNKAGTFETKKTLPKLAQAAGNADAWASSPOSGSGGATTVSDSGSSGSDST	428
Db	155	--TASSVGVTEAR---PSVAGSGTTGTV-----SGA---SGSTSSSGSPGAT	195
Qy	429	SETV--PVYAKGGGLYTDKNLSITNITGIIETANNKATDVGGGAYVKGTLTCENSHRLQF	486
Db	196	GASIGQPETSR-----ISVAGSSG-----APAVSSGA-----222	
Qy	487	LKNSSDKGGGTYGEDNITLSNLTGKTLFQENTAKEGGGLFIKGTDKALTMGLDLSFCL	546
Db	223	-----SQAAGTSGAGPGTTASSVGVTEARPS-----VAGSGTGTVSGAS---263	
Qy	547	INNTSEKHHGGAFVTKETISQYTTSDVETIPGTPVHGETVIITGNKS-----TGNNGGGVCT	602
Db	264	-GSTSSSGSPGATGASIGQPETSRIS---SVAGSSGAPAVSSGASQAAGTSGAGPGTTA	318
Qy	603	KRLALSNLQSIISICNSAAENGGAHTCPDSFPTADTAEQPAASAAATSTPKSAVPSTAL	662
Db	319	SSVGVTEARPSVAGSGTTGTVSGA-----SGSTSSSGSPGATGASIGQPETSRISV	373



**THIS PAGE BLANK (USPTO)**



PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
PT comprises immunogenic portion of Chlamydia antigen, which comprises  
PT amino acid sequence encoded by polynucleotide sequence -  
XX  
PS Claim 1; Pages 177-179; 256pp; English.  
XX  
CC The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamidiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and can  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a nucleic acid sequence  
CC isolated in the present invention.  
XX  
SQ Sequence 5265 BP; 1625 A; 1207 C; 1023 G; 1410 T; 0 other;

Query Match 100.0%; Score 5265; DB 21; Length 5265;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gcaatcatgaaatgctgcagctactcgggtgttctgtgtctgtctccctcagtttca 60  
Db 1 gcaatcatgaaatgctgcagctactcgggtgttctgtgtctgtctccctcagtttca 60  
Qy 61 ggggtttgtctccagaaacctaagaatttaattctctcgtctagaactttctcct 120  
Db 61 ggggtttgtctccagaaacctaagaatttaattctctcgtctagaactttctcct 120  
Qy 121 accacttttactgaacaattggagaagctgggcagaaatatatcgtctcgttaacgca 180  
Db 121 accacttttactgaacaattggagaagctgggcagaaatatatcgtctcgttaacgca 180  
Qy 181 tctttcacaaaatttacaaacattctactaccgatacaacaactcccacgaactcaaac 240  
Db 181 tctttcacaaaatttacaaacattctactaccgatacaacaactcccacgaactcaaac 240  
Qy 241 tctctagtctcgcggagaactctccgtttctcgtgagatagtgactctacaacaacg 300  
Db 241 tctctagtctcgcggagaactctccgtttctcgtgagatagtgactctacaacaacg 300  
Qy 301 actcctgatcctaagggtggcggcccttttataacgcgcactccggagtttctccttt 360  
Db 301 actcctgatcctaagggtggcggcccttttataacgcgcactccggagtttctccttt 360  
Qy 361 atgacacatcaggaaacagaaggttccttaactcgtcgtgagataaaatgactgtgaa 420  
Db 361 atgacacatcaggaaacagaaggttccttaactcgtcgtgagataaaatgactgtgaa 420  
Qy 421 ggcggtgtctattctctcaaggagagctgtctattacagatctcacagttcaaccatc 480  
Db 421 ggcggtgtctattctctcaaggagagctgtctattacagatctcacagttcaaccatc 480  
Qy 481 caaaaataactatccagctatccggaggagcgatttttgaggatctacaatctcccta 540  
Db 481 caaaaataactatccagctatccggaggagcgatttttgaggatctacaatctcccta 540  
Qy 541 tcaggattactaaagcgaactttctcgtcaactcgcagaagttccctgctcctttaag 600  
Db 541 tcaggattactaaagcgaactttctcgtcaactcgcagaagttccctgctcctttaag 600  
Qy 601 aaacctacagaacctaaagctcaaacagcgaacgctcgtggttcttagttctctagc 660  
Db 601 aaacctacagaacctaaagctcaaacagcgaacgctcgtggttcttagttctctagc 660  
Qy 661 ggaatatgattcgtgtcttccccccagttccacgtatagatgaaacccgcgcagctaatctt 720

Db 661 ggaatatgattcgtgtcttccccccagttccacgtatagatgaaacccgcgcagctaatctt 720  
Qy 721 caaagtcactttatttggctacagctactcgtcgtctcaaacgcgatacacgaaacatca 780  
Db 721 caaagtcactttatttggctacagctactcgtcgtctcaaacgcgatacacgaaacatca 780  
Qy 781 actcctctcaaaagccagagctcggggagctactctgctaaagcgacacttactatc 840  
Db 781 actcctctcaaaagccagagctcggggagctactctgctaaagcgacacttactatc 840  
Qy 841 gcagactctcaaggggtactattctcaataataaagctactaaagatggaggagcgatc 900  
Db 841 gcagactctcaaggggtactattctcaataataaagctactaaagatggaggagcgatc 900  
Qy 901 tttgctgaaagatgtttcttctcgagaattattacattataaaagtaacaaactaacgt 960  
Db 901 tttgctgaaagatgtttcttctcgagaattattacattataaaagtaacaaactaacgt 960  
Qy 961 gctgaagaaaaggagagctatctatgctaaagtgacctctcaattcaattcttctaaa 1020  
Db 961 gctgaagaaaaggagagctatctatgctaaagtgacctctcaattcaattcttctaaa 1020  
Qy 1021 cagagctcttttaattctaaactacagtaacaagtggggggctctctatatgttgaaagga 1080  
Db 1021 cagagctcttttaattctaaactacagtaacaagtggggggctctctatatgttgaaagga 1080  
Qy 1081 ggtataaaacttccaagatctctgaaagaaattcgcattaaagtaacataaaagctggaacgttc 1140  
Db 1081 ggtataaaacttccaagatctctgaaagaaattcgcattaaagtaacataaaagctggaacgttc 1140  
Qy 1141 gaaacaaaaaaacacatttaccctctttaaagctcagcatctgcaggaaatgcagat 1200  
Db 1141 gaaacaaaaaaacacatttaccctctttaaagctcagcatctgcaggaaatgcagat 1200  
Qy 1201 gcttgggctcttctcctctccaaatctggtctgagcaactacagtgctccgactcagga 1260  
Db 1201 gcttgggctcttctcctctccaaatctggtctgagcaactacagtgctccgactcagga 1260  
Qy 1261 gactctagctctgctcagactcggatcacctcagaaacagttccagtcacagctaaagc 1320  
Db 1261 gactctagctctgctcagactcggatcacctcagaaacagttccagtcacagctaaagc 1320  
Qy 1321 ggtgggcttatactgataaagaattcttcgattactaacatcacaggaattatcgaatt 1380  
Db 1321 ggtgggcttatactgataaagaattcttcgattactaacatcacaggaattatcgaatt 1380  
Qy 1381 gcaataacaaagcgacagatgttgagggtggtcgtacgtataaaaggaacccctactgt 1440  
Db 1381 gcaataacaaagcgacagatgttgagggtggtcgtacgtataaaaggaacccctactgt 1440  
Qy 1441 gaaaactctcacgctctacaatttttgaanaactctccgataaaacaggtggaggaaatc 1500  
Db 1441 gaaaactctcacgctctacaatttttgaanaactctccgataaaacaggtggaggaaatc 1500  
Qy 1501 tacggagaagaacaactcacccctatctaatgtgacaggaagactctatcccaagagaat 1560  
Db 1501 tacggagaagaacaactcacccctatctaatgtgacaggaagactctatcccaagagaat 1560  
Qy 1561 actgccaagaagagggcgggtgagctcttcaataaagggtacagaataaagctcttacaatg 1620  
Db 1561 actgccaagaagagggcgggtgagctcttcaataaagggtacagaataaagctcttacaatg 1620  
Qy 1621 acaggactggatgttctgttttaataacacatcagaaaaacatgggtggaaccc 1680  
Db 1621 acaggactggatgttctgttttaataacacatcagaaaaacatgggtggaaccc 1680  
Qy 1681 tttgttaccaaaagaatactctcagacttacacctctgatgtggaacaaattccaggaatc 1740  
Db 1681 tttgttaccaaaagaatactctcagacttacacctctgatgtggaacaaattccaggaatc 1740  
Qy 1741 acgctctgacatggcgaaacagcttactctggcaataaattacagggaggttaattggtgga 1800

Db 1741 acgcgttacatggtgaacagtcattactggtgcaataaaattcacaggaggaataatggtgga 1800  
QY 1801 ggcggtgtacaaaacgcttgccttatctaaaccttaacgttcaaaagcatttctatatccgggaat 1860  
Db 1801 ggcggtgtacaaaacgcttgccttatctaaaccttaaaagcatttctatatccgggaat 1860  
QY 1861 tctgcagcagaaaaatggtggtgagcccccacacatgccagatagcttcccaacgcggaat 1920  
Db 1861 tctgcagcagaaaaatggtggtgagcccccacacatgccagatagcttcccaacgcggaat 1920  
QY 1921 actgcagaacagccgcagcagcttctgcgcgacgtctactctactcccaaatctgcccgggtc 1980  
Db 1921 actgcagaacagccgcagcagcttctgcgcgacgtctactctactcccaaatctgcccgggtc 1980  
QY 1981 tcaactgcttaagcacaccttcatctctacogtctcttcatctaaccttaactagcagcc 2040  
Db 1981 tcaactgcttaagcacaccttcatctctacogtctcttcatctaaccttaactagcagcc 2040  
QY 2041 tcttcacaagcctctcctgcgaacctctaaataaggaaactcaagaatcctaatgctgtataca 2100  
Db 2041 tcttcacaagcctctcctgcgaacctctaaataaggaaactcaagaatcctaatgctgtataca 2100  
QY 2101 gacttattgaatcgattatgattagttatagactatcagcaaaaacactgctaagaagcc 2160  
Db 2101 gacttattgaatcgattatgattagttatagactatcagcaaaaacactgctaagaagcc 2160  
QY 2161 ggtggaaatctatgtctaaaaagccaaagatgtcccgcataagacaaactgataatctctgag 2220  
Db 2161 ggtggaaatctatgtctaaaaagccaaagatgtcccgcataagacaaactgataatctctgag 2220  
QY 2221 aactcgcgtacagagataggtgaggtatctgctgttaaagaaatcttttagaactagatgct 2280  
Db 2221 aactcgcgtacagagataggtgaggtatctgctgttaaagaaatcttttagaactagatgct 2280  
QY 2281 ctagtctccttactgttaacagagaacctgtgtggaaagaaggtggaggtctacatgct 2340  
Db 2281 ctagtctccttactgttaacagagaacctgtgtggaaagaaggtggaggtctacatgct 2340  
QY 2341 aaaaactgtaaatatttctaacttgaatcaggcttctcttctcgacaacaaagcaaac 2400  
Db 2341 aaaaactgtaaatatttctaacttgaatcaggcttctcttctcgacaacaaagcaaac 2400  
QY 2401 tctctatccacagaggtgcgaacaaacagcttcagcactgctgcagctgctgtcttcccta 2460  
Db 2401 tctctatccacagaggtgcgaacaaacagcttcagcactgctgcagctgctgtcttcccta 2460  
QY 2461 caagcagccgacgacgcaccatcatctccagcaacccaacttattcaggtgtagta 2520  
Db 2461 caagcagccgacgacgcaccatcatctccagcaacccaacttattcaggtgtagta 2520  
QY 2521 ggaggagctatctatggagaaaagggttacattctctcaatgtagcgggacttgcagttc 2580  
Db 2521 ggaggagctatctatggagaaaagggttacattctctcaatgtagcgggacttgcagttc 2580  
QY 2581 tctgggaacaaagctatcgataaacaatccctcccaatcatcgtgtgaacgtacaaggagga 2640  
Db 2581 tctgggaacaaagctatcgataaacaatccctcccaatcatcgtgtgaacgtacaaggagga 2640  
QY 2641 gccatctatgcgaacaaaccttctgtctatttggatctctcgatctggaacctcttatatt 2700  
Db 2641 gccatctatgcgaacaaaccttctgtctatttggatctctcgatctggaacctcttatatt 2700  
QY 2701 ttctcggggaacagtgcttccactgggaaatctcaaacacaggggcaaatagcgggagga 2760  
Db 2701 ttctcggggaacagtgcttccactgggaaatctcaaacacaggggcaaatagcgggagga 2760  
QY 2761 ggaatctatccctactgcttaacattgaattgtccctgcgacatactcttaacaatacagcc 2820  
Db 2761 ggaatctatccctactgcttaacattgaattgtccctgcgacatactcttaacaatacagcc 2820  
QY 2821 tctatagctacaccggaagacttctctggaatggatctcctcaggaaatctctattaaagat 2880  
Db 2821 tctatagctacaccggaagacttctctggaatggatctcctcaggaaatctctattaaagat 2880

QY 2881 accattggagagccattgcaggagcagccattaccctatctctgagctctctcgattttca 2940  
Db 2881 accattggagagccattgcaggagcagccattaccctatctctgagctctctcgattttca 2940  
QY 2941 gggaaatacggctgatttaggagctgcaataggaaactctagctaatgcaaatacaccagtt 3000  
Db 2941 gggaaatacggctgatttaggagctgcaataggaaactctagctaatgcaaatacaccagtt 3000  
QY 3001 gcaactagcggatctcaaaatagcattacagaaaaaattacttttagaaaaacggtctcttt 3060  
Db 3001 gcaactagcggatctcaaaatagcattacagaaaaaattacttttagaaaaacggtctcttt 3060  
QY 3061 atttttgaagaaacaaagctaataaacgttggagcgatttactctctctagcgtttccatt 3120  
Db 3061 atttttgaagaaacaaagctaataaacgttggagcgatttactctctctagcgtttccatt 3120  
QY 3121 aaagggaataataattaccttcaatcaaaaaatcacatccactcatgatggaagcgtctatctac 3180  
Db 3121 aaagggaataataattaccttcaatcaaaaaatcacatccactcatgatggaagcgtctatctac 3180  
QY 3181 ttacaaaaagatcgctacgattgagcttcttttagatctgttcttttttacagaaaaatacgtt 3240  
Db 3181 ttacaaaaagatcgctacgattgagcttcttttagatctgttcttttttacagaaaaatacgtt 3240  
QY 3241 acagctacacaagctagtctgcaacatctggcaaaatacacaataactgccaactatggg 3300  
Db 3241 acagctacacaagctagtctgcaacatctggcaaaatacacaataactgccaactatggg 3300  
QY 3301 gcaagcaatctttggaagatccaggaaacacactcaatcgctcaaacagatgcaattttaacc 3360  
Db 3301 gcaagcaatctttggaagatccaggaaacacactcaatcgctcaaacagatgcaattttaacc 3360  
QY 3361 ctctctgcttctctggaaaacattacttttagcaaaacagtttacagaataaccaaggt 3420  
Db 3361 ctctctgcttctctggaaaacattacttttagcaaaacagtttacagaataaccaaggt 3420  
QY 3421 gatactccgcgtcagcaagtttttagtattgcagggaacgctcaaaactctctctacaagcc 3480  
Db 3421 gatactccgcgtcagcaagtttttagtattgcagggaacgctcaaaactctctctacaagcc 3480  
QY 3481 gctaaaggggaagactatttagcttttttgcattgtgtgcacacctctacccaaaaaacggt 3540  
Db 3481 gctaaaggggaagactatttagcttttttgcattgtgtgcacacctctacccaaaaaacggt 3540  
QY 3541 tcaacacaaaacgtttatgaaacttttagatatttaataaagaacagacagatatacctat 3600  
Db 3541 tcaacacaaaacgtttatgaaacttttagatatttaataaagaacagacagatatacctat 3600  
QY 3601 acaggaaactattgttctctctctgaattacatgaaacaaaatcttacatccccacagaat 3660  
Db 3601 acaggaaactattgttctctctctgaattacatgaaacaaaatcttacatccccacagaat 3660  
QY 3661 gcaatctctcaaacggaaacttttagttctttaaagagaaaaacagactccacgtagtctct 3720  
Db 3661 gcaatctctcaaacggaaacttttagttctttaaagagaaaaacagactccacgtagtctct 3720  
QY 3721 tttagcagaaagaaggtctaaaatttaattatggaacccggagctgtgttatctaaccacaa 3780  
Db 3721 tttagcagaaagaaggtctaaaatttaattatggaacccggagctgtgttatctaaccacaa 3780  
QY 3781 aacatagctaacgcgagctcttagctatcaatgggttaacgattgatctttccagtatgggg 3840  
Db 3781 aacatagctaacgcgagctcttagctatcaatgggttaacgattgatctttccagtatgggg 3840  
QY 3841 actcctcaagcaggggaaaaatctctctctccagaaattacgtatcgttgcacgaacctct 3900  
Db 3841 actcctcaagcaggggaaaaatctctctctccagaaattacgtatcgttgcacgaacctct 3900  
QY 3901 agtgcacatccgaggaagcggggtcagcagtagtatataccaacaaatccttaaaagatttct 3960  
Db 3901 agtgcacatccgaggaagcggggtcagcagtagtatataccaacaaatccttaaaagatttct 3960

QY 3961 ggcagcgtcctcagggttcgcccgaactactcgaactatgagcgagaaacaaagtcttc 4020  
Db |||||||  
QY 3961 ggcagcgtcctcagggttcgcccgaactactcgaactatgagcgagaaacaaagtcttc 4020  
Db |||||||  
QY 4021 ctacacagagaccttactttaaataagatcctaataagaaacttttaccacaaacctatggtta 4080  
Db |||||||  
QY 4021 ctacacagagaccttactttaaataagatcctaataagaaacttttaccacaaacctatggtta 4080  
Db |||||||  
QY 4081 ggaagcgatcagatgtaccactaataagcttcgacttaacacaaagtgagcgtccaaagt 4140  
Db |||||||  
QY 4081 ggaagcgatcagatgtaccactaataagcttcgacttaacacaaagtgagcgtccaaagt 4140  
Db |||||||  
QY 4141 tatgatttacttctcggggtcttttccctcagaaaggtacatggtgaacctgggaca 4200  
Db |||||||  
QY 4141 tatgatttacttctcggggtcttttccctcagaaaggtacatggtgaacctgggaca 4200  
Db |||||||  
QY 4201 ttagatttacttccacaaacagggaacttcaagccagatggacattcgatcacctatcgt 4260  
Db |||||||  
QY 4261 cgtcgggtatcacatccctaggatcaatttttatcgaactctatcttaggtcccca 4320  
Db |||||||  
QY 4261 cgtcgggtatcacatccctaggatcaatttttatcgaactctatcttaggtcccca 4320  
Db |||||||  
QY 4321 aactcaatgattgttgaagcaagggtcttacaacaacatgttgaataatgcccgcttc 4380  
Db |||||||  
QY 4321 aactcaatgattgttgaagcaagggtcttacaacaacatgttgaataatgcccgcttc 4380  
Db |||||||  
QY 4381 gatgatcgtcttacaataactctcgggtttcagagtaggaactttctttagctcaacaa 4440  
Db |||||||  
QY 4381 gatgatcgtcttacaataactctcgggtttcagagtaggaactttctttagctcaacaa 4440  
Db |||||||  
QY 4441 ggaactcctcttccgaagaattcagttactacagccgcggaacttcaagtgcacatgat 4500  
Db |||||||  
QY 4441 ggaactcctcttccgaagaattcagttactacagccgcggaacttcaagtgcacatgat 4500  
Db |||||||  
QY 4501 gccaaacctagacaagattttatcctagagctgcatttagtaagatagtggtggaaaacc 4560  
Db |||||||  
QY 4501 gccaaacctagacaagattttatcctagagctgcatttagtaagatagtggtggaaaacc 4560  
Db |||||||  
QY 4561 aaagccatacaaaaaatgcataacttacttccataaaggcctcgtgaactcttaccacaa 4620  
Db |||||||  
QY 4561 aaagccatacaaaaaatgcataacttacttccataaaggcctcgtgaactcttaccacaa 4620  
Db |||||||  
QY 4621 tctgtctatggaggtcaaatctcgtatttcttgcctcaataaagcaacatggttggcactt 4680  
Db |||||||  
QY 4621 tctgtctatggaggtcaaatctcgtatttcttgcctcaataaagcaacatggttggcactt 4680  
Db |||||||  
QY 4681 cctttcctaatacaaggagctgtcctatggacatattaaacatgatacaacaacactt 4740  
Db |||||||  
QY 4681 cctttcctaatacaaggagctgtcctatggacatattaaacatgatacaacaacactt 4740  
Db |||||||  
QY 4741 taccctctcctatcgaagaataaagagagattgggaagatttaggattggttagcgat 4800  
Db |||||||  
QY 4741 taccctctcctatcgaagaataaagagagattgggaagatttaggattggttagcgat 4800  
Db |||||||  
QY 4801 ctctgtatctctatgattctttaaagaaccttcaagattcttctaaacggatcactgtc 4860  
Db |||||||  
QY 4801 ctctgtatctctatgattctttaaagaaccttcaagattcttctaaacggatcactgtc 4860  
Db |||||||  
QY 4861 tatggggaactcagatattccagcattccgcagaaacagttcacagaaatcgattacgat 4920  
Db |||||||  
QY 4861 tatggggaactcagatattccagcattccgcagaaacagttcacagaaatcgattacgat 4920  
Db |||||||  
QY 4921 ccaagacacttcagatgattgtcttcaagaaatctcgtcctcctcgtgggatcgctgtc 4980  
Db |||||||  
QY 4921 ccaagacacttcagatgattgtcttcaagaaatctcgtcctcctcgtgggatcgctgtc 4980  
Db |||||||  
QY 4981 gaaggagctatcatgaactgtaattcttctatgataataaagcttgcattagcctcacatg 5040  
Db |||||||  
QY 4981 gaaggagctatcatgaactgtaattcttctatgataataaagcttgcattagcctcacatg 5040  
Db |||||||  
QY 5041 ccttctatctacagaaataactcgtctgtgtaataatcgggtattgtcttcgaatgaagct 5100

Db |||||||  
QY 5041 ccttctatctacagaaataactcgtctgtgtaataatcgggtattgtcttcgaatgaagct 5100  
Db |||||||  
QY 5101 ggtcaagttatctcggagtagtccaactagaacctctgtagacagaataacagtactcaa 5160  
Db |||||||  
QY 5101 ggtcaagttatctcggagtagtccaactagaacctctgtagacagaataacagtactcaa 5160  
Db |||||||  
QY 5161 ctatatctcgtcctcctcgtgactctctacgaaactatactatcgtatgtaggcatgtat 5220  
Db |||||||  
QY 5161 ctatatctcgtcctcctcgtgactctctacgaaactatactatcgtatgtaggcatgtat 5220  
Db |||||||  
QY 5221 acgctatcgcaaatgactagctggtgctgcgcgatgtcttcttaa 5265  
Db |||||||  
QY 5221 acgctatcgcaaatgactagctggtgctgcgcgatgtcttcttaa 5265  
Db |||||||

RESULT 2

AAZ01425/c  
ID AAZ01425 standard; DNA; 1038602 BP.

XX AC AAZ01425;

XX DT 07-OCT-1999 (first entry)

XX Complete genome sequence of Chlamydia trachomatis.

XX Vaccine: eye disease; conventional trachoma; nonendemic trachoma;  
XX paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
XX nongonococcal urethritis; epididymitis; salpingitis;  
XX bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

OS Chlamydia trachomatis.

XX WO9928475-A2.

XX PD 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077

PR 28-NOV-1997; 97US-0095041

PR 17-DEC-1997; 97US-0095041

XX (GEST ) GENSET.

XX Griffiths R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

PS Claim 1; Page 373-656; 1755pp; English.

XX The present sequence represents the complete genome of Chlamydia  
XX trachomatis. Open reading frames (ORFs) of the genome encode  
XX polypeptides AA376754-Y37949. The polypeptides can be used as vaccines  
XX against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
XX be used to control growth of the microorganism. Chlamydia trachomatis is  
XX responsible for a large number of diseases, e.g. eye diseases such as  
XX conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion  
XX conjunctivitis; genital diseases such as nongonococcal urethritis,  
XX epididymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;  
XX pneumopathy in breast feeding infants; and venereal  
XX lymphogranulomatosis. The polypeptides of the invention may be of use in  
XX treating these diseases.

SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match

Best Local Similarity 95.0%; Score 4999.2; DB 20; Length 1038602;

Matches 5107; Conservative 0; Mismatches 74; Indels 17; Gaps 5;

Qy 72 ccagaaaccttaagaattaaattctctcgcgtagaaaacttcttcctctaccacttttac 131  
Db 1038602 CCAGAACTTANNNNNTAAATTTCTCTCGGTAGGAACCTTCTTCGCTACCACACTTTTTAC 1038543

Qy 132 tgaacaattgagaacatggcgagaaatatatcgctctctgtgtaacgcatctttcacaaa 191  
Db 1038542 TGAACACNGTTGGAGAAGCTGGGCGAANTATATCGTCTCTAGTAAGCAATCTTTCACAAA 1038483

Qy 192 attaccaaacattcctactaccgatatacaaacctccacgaactccaaactcctctagtct 251  
Db 1038482 ATTTACCAACATTCCCTACTACCAATAACAACAACCTCCCACGAACCTCAAACTCCTTAGCTC 1038423

Qy 252 tagcggagaaactgcttcctgtcttgaggatagtagtactctatacaacaacgacctcgtatcc 311  
Db 1038422 TAACAGAGACACTGCTTTCCGCTTTCTAAGGATAGTGACTCTACAAACAGCACTCCTGATCC 1038363

Qy 312 taaagtgcgcgcccttttataacgcgcacactccggagttttgccitttatgacacatc 371  
Db 1038362 TAAAGTGGCGCGGCNFTTTATAACGGCACTCCGGAGTTTGTCTTTATGACACGATC 1038303

Qy 372 aggaacgaaggttctctaactctgtctgagataaaaaatgactggtgaagcgqctgctat 431  
Db 1038302 AGGAACAGAAAGTTCCCTTAACCTGTCTGAGATAAAAAAGACTGGTGAAGCGCGTGCTAT 1038243

Qy 432 ctctctcaaggagagctgctatttacagatcttgacaagtttaaccattccaataaactt 491  
Db 1038242 CTTCTCTCAAGGAGAGCTGCTATTTTACAGATCTGACAGGTCTTAACCATCCAAAATAAATTT 1038183

Qy 492 atccagctatccgaggagcgatttttgaggatctacaactcccctatcaggagattac 551  
Db 1038182 ATCCCAGCTATCCGAGGAGCGAATTTTGGAGATCTTACAAATCTCCTATPCAGGGGATTTAC 1038123

Qy 552 taaagcgacttctctgcaactctgcaagagtttctctgctctctgtttaagaacaacctacaga 611  
Db 1038122 TAAAGCGACTTTCTCCTCCAACCTGCGAAGTTCTCTGCTCCTGTTAAGAAACCTTACAAA 1038063

Qy 612 acctaaagctcaaacagcaagcgaacgtaacgctctgtagttctagtcgagaaatgattc 671  
Db 1038062 ACCTGAAGCTCAAACAGCAGCGCAAACGTCGGGTCTTAGTGTCTTAGCGGAATGATTC 1038003

Qy 672 ggtgtctccccagttccagtagagctgaaccgcgacgactaatctcaaatgacttt 731  
Db 1038002 GGGCTCTTCTCCAGTTCAGTTCCAGTAGTAAGCTGAACCGCACGAGCTAATCTTCAAAAGTCAC TT 1037943

Qy 732 tatttgtctacagctactcctgctgctcctcaaaccgatatacagaacaacatacctccctctca 791  
Db 1037942 TATTGTGTGCTACAGCTACTCTCTGCTCAACCGGATACAGAAACATCAACTCCCTCTCA 1037883

Qy 792 taagccaggatctctggggagctatctatgctaaaggcgacctactactcgagactctca 851  
Db 1037882 TAAAGCCAGGATCTGGGGAGCTATCTATGCTAAAGCGCACCTTTACTATCGACACTCTCA 1037823

Qy 852 agaagttactattccaataaaataaagctactaaagatggaggagcgatcttttgtgagaa 911  
Db 1037822 AAAGGTACTATTCTCAATAAATAAAGCTACTAAAGATGGAGGAGCGATCTTCTCCTCAGAA 1037763

Qy 912 agatgtttcttcgagaatattacatcataaaagtcacaaactaacgggtgctgagaaaa 971  
Db 1037762 AGATGTTTCTTCGAGNAATATTACATCATTTAAAGTACAAACTTAACGGTGCTGAAGAAA 1037703

Qy 972 gggaggagctatctatgctaaaggtagcctctcaactctcaactcttcttaaacagagctctttt 1031  
Db 1037702 GGGAGGAGCTATCTATGCTAAAGGTGACCTCTCAATTTCAATCTTCTAAACAGAGTCTTTT 1037643

Qy 1032 taattctaactacagtaaaacaagggtggggggctctatatgttgaaaggagataaactt 1091  
Db 1037642 TAAATTTCACTACAGTAACAACAGGTGGTGGGGCTCTATATGTTTGAAGGAATAATAA ACTT 1037583

Qy 1092 ccaagatcttgaagaattctgcatataagtaacaataaagcttggaacggttcgaacaacaaaa 1151  
Db 1037582 CCAAGATCTTTGAAGAAATTCGCATTAATTAAGTACAATAAAGCTGGAAACGTTCAAAACAAAAA 1037523

Qy 1152 aatcactttactcttttaaaagctcaagcatctcgaggaaatgcagatgcttgggctc 1211

Dbb 1036445 CTACAGAGATAGGTGGAGGTATCTGCTGTAAAAAATCTTTAAAAACTAGATGCTCTAGTCT 1036386  
QY 2288 ccttatctgtaacagagaaacctgttgtggaagaaggtggaggttacatctaaacctg 2347  
Dbb 1036385 CCTTATCTGTAACAGAGAACTTGTGTGGGAAGAAGGTGGAGGCTTACATGCTAAAACTG 1036326  
QY 2348 taaatatttctaattgaaatcagagcttctcttctcgaacaaacaaagcaaaactctcat 2407  
Dbb 1036325 TAAATATTTCTAATCTGAATCAGGCTTCTTTCTCGAACAAACAAAGCAAACTCCCTCAT 1036266  
QY 2408 ccaagagggtcgcaacaacagcttcaagcaactgtctgagctgtctcctccataacagag 2467  
Dbb 1036265 CCACAGAGGTGCGAACACAGCTTTCACACAGCTGTGAGCTGTCTCCCTCAACAGAG 1036206  
QY 2468 ccgagagccgcacccatcatctccagcaacacacacattatcaggtgtagtagagagag 2527  
Dbb 1036205 CCGCAGCAGCGCCACCATCATCTCCAGCAACACCACTTATTCAGGTGTAGTAGGAGGAG 1036146  
QY 2528 ctatctatggagaaaggttacattctctcaatgttagcgggacctgtgcagttctctctggga 2587  
Dbb 1036145 CTATCTATGGAGAACAGTTTACATTCTCTCAATGTACGGGACTTGTCTCAGTTCTCTGGGA 1036086  
QY 2588 accaagctatcgataacaatccctcccaatcatctgttgaacgtacaaagagagccatct 2647  
Dbb 1036085 ACCAAGCTATCGATAACAATCCCTCCCAATCATCGTTGAACGTACAAAGGAGGCCATCT 1036026  
QY 2648 atgccaaaaccttctgtattggatcttccgagcttccgagcttccatatttttctcgg 2707  
Dbb 1036025 ATGCCAAAACCTTTGTCTATTAGATCTTCCGATGCTGGAACCTCTATATTTTCTCGG 1035966  
QY 2708 ggaacagtgctccactgggaaatctcaaaacagggcaaatagcgggagcgatct 2767  
Dbb 1035965 GGAACAGTGTCTCAGTGGGAATCTCAACACAGGGCAAAATAGCGGAGGACGATCT 1035906  
QY 2768 actccctactgtttacatgtgaattgcctgagcaattctctaaacaatcagcctctatag 2827  
Dbb 1035905 ACTCCCTACTGTGTACATTGAAATTTGCTGCGACATTTCTTAACAATACAGCCTCTATGG 1035846  
QY 2828 ctacaccgaagctctctcgaagatgatcctcaggaattctattaaagataccattg 2887  
Dbb 1035845 CTACACCGAAGACTTCTC-----ATCCTCAGGAAATCTATTAAAGATACCAATG 1035795  
QY 2888 gagggaccattgagggagagccattaccctatctggtgagctctcgtatttccagggaata 2947  
Dbb 1035794 GAGGAGCCATTGACGGGACAGCCATTACCCTATCTGAGTCTCTCGATTTTCAGGGGAATA 1035735  
QY 2948 cggctgattaggagcgtcaataggaaactctagctaaatgcgaatacacccagtgcaacta 3007  
Dbb 1035734 CGGCTGATTTAGGAGCTGCAATAGGAACTCTAGCTAATGCAAAATACACCCAGAGCAACTA 1035675  
QY 3008 gcggatctcaaaatagcattacagaaaaaattactttagaaaaaggttctttattttg 3067  
Dbb 1035674 GCGGATCTCAAAATAGCATTTACAGAAAAAATTACTTTAAAAACAGTCTTTTATTTTGG 1035615  
QY 3068 aaagaaaaagcgttaataaacgtggagcgtttactctctcagcgtttccattaaaggga 3127  
Dbb 1035614 AAAGAAAACCAAGCTAATAAACGTGGAGCGATTTACTCTCTAGCGTTTCCATTAAAGGGA 1035555  
QY 3128 ataattattacctcaatacaaatacatccatcatgtatggagcgtatctactttcaaa 3187  
Dbb 1035554 ATAATATTTACCTTCAATCAAAATACATCCACCTCATGATGGAAGCGCTATCTACTTTTACAA 1035495  
QY 3188 aagatgctcacgattgagctttagatctgttctttttacaggaataacgcttacagcta 3247  
Dbb 1035494 AAGATGCTAAGATTGAGTCTTTTAGATCTGTCTTTTACAGGAAATTAAGATTACAGCTA 1035435  
QY 3248 cacaagctagttctgcaacatctggacaaaaatacaaaatctatgccaactatggggcagcca 3307  
Dbb 1035434 CACAAGCTAGTTCTGCAACATCTGCACAAAATACAAATACTGCCAACTATGTGGGCAGCTA 1035375  
QY 3308 totttggagatccagaaccactcaatcgtctcaaaacagatgcgaattttaacccttcttg 3367  
Dbb 1035374 TCCTTTGGAGATCCAGGAACCACTCAATCGCGCTCAACAGATGCCATTTTAAACCCCTTTG 1035315

QY 3368 ctctcttggaacattacttcttagcaacaacagtttacagaataaccaaggtgatactc 3427  
Dbb 1035314 CTCTCTCTGGAACATTTACTTTTAGCAACAACAGTTTTACAGATAACCAAGGTGGTACTC 1035255  
QY 3428 ccgctagcaagttttgttagtattgcagagatacgtcaaaactctctcaaacgcgcftaaag 3487  
Dbb 1035254 CCGCTAGCAAGTTTGTGTAGTATTGCAGGATACGTCAAACTCTCTTTTACAAAGCCCAATAAG 1035195  
QY 3488 ggaagactattagctttttcgtattgtgtgcacacctctacaaaaaaacaggtttcaaacac 3547  
Dbb 1035194 GGAAGACTATTAGCTTTTTCGATTGTGTGCACACCTCTACCAAAAAACAGGTTTCAACAC 1035135  
QY 3548 aaaaactttatgaaacttttagatatataaagaagaagaaacagtaataatccatatacaggaa 3607  
Dbb 1035134 AAAACGTTTATGAACTTTTAGATATTAAATAAAGAGAGACAGTAAGCCATATACAGGAA 1035075  
QY 3608 ctattgttctctctctgaattacatgaaaaaaatcttatcccccacagaatgcaatcc 3667  
Dbb 1035074 CTATTGTGTCTCTCTGTAATTACATGAAAAAATACTTACATCCACAGAAATGCAATCC 1035015  
QY 3668 ttcacaacggaacttttagtttcttaaaagaaaaacagaactccacgtagtctcttttggagc 3727  
Dbb 1035014 TTCACAACGGAACTTTTAGTTCTTAAAGAGAAAAACAGAACTCCACGTAGTCTCTTTTGAGC 1034955  
QY 3728 agaaagaagggtctaaaaatgaattatggaaacccggagctgtgttatctaaacaaaacatacg 3787  
Dbb 1034954 AGAAAGAAAGGCTCTAAATTAATTAATGGAACCCGGAGCTGTGTTATCTTAACCAAAACATAG 1034895  
QY 3788 ctaacggagctctagctatcaatgggttaacgattgatcttccagtagtgggagctctc 3847  
Dbb 1034894 CTAAACGGAGCTTAGCTATCAATGGGTTAAAGATTGATCTTTTCCAGTTATGGGACTCCTC 1034835  
QY 3848 aagcaagggaaaacttctctctccagaaattacgtatogttgcacagacctctagtgat 3907  
Dbb 1034834 AAGCAGGGGAAATCTCTCTCCCTCCAGAAATACGATATCGTTGCCACACCTCTAGTGCAT 1034775  
QY 3908 ccgaggaagcgggctcagcagtagtataaccaaaaaatctctaaaggatttctcgcagcag 3967  
Dbb 1034774 CCGGAAGAAAGCGGGTCAGCAGGAGTATACCAACAATCTTAAGGGATTTCTGCAGCAG 1034715  
QY 3968 tgccttcagggtctgcgcgaactactccaactatgagcgagaacaaagtcttctcctaaacag 4027  
Dbb 1034714 CGCCTTCAGGTTCTGCGCGCAACTACTCCAATATGAGGAGAGAACAAAAGTTTTCCTAACAG 1034655  
QY 4028 gagaccttactttaatacactcctaagaaacttttaccaaaacctatgtagaaagcg 4087  
Dbb 1034654 GAGACCTTACTTTAATAGATCCTAATGGAACCTTTTATCAAAACCCCTATGTTAGGAAGCG 1034595  
QY 4088 atctagatgtaccactaataatgaagcttccgacttaacacaagtgacgtccaagtctatgatt 4147  
Dbb 1034594 ATCTAGATGTACCCTAATTAAGCTTCCGACTAACACAAGTGACGTCCAAGTCTATGATT 1034535  
QY 4148 taactttatctgggagatcttttccctcagaaagggtacatgggaacctggcaatttagatt 4207  
Dbb 1034534 TAACCTTATCTGGGATCTTTTCCCTCAGAAAGGGTACATGGGAACCTGGACATTAATTT 1034475  
QY 4208 ctaatcccaaacagggaaaacttcaagcagatgaaactatcgatcctatcgctcgtggg 4267  
Dbb 1034474 CTAATCCACAACAGGAAACTTCAAGCCAGATGGACATTGATACCTATCTCGTCGCTGGG 1034415  
QY 4268 tatcacatccctagggaataatcatttttatggaaactctatcttaggctcccaaaaactcaa 4327  
Dbb 1034414 TATACATACCTAGGGATTAATCATTTTATGGAACTCTATCTTAGGCTCCCAAACTCAA 1034355  
QY 4328 tgaattgtgtgaagcaagggtcttatcaacaacatgttgtaataatgcccgcttcgatgata 4387  
Dbb 1034354 TGATTGTGTGAAGCAAGGGCTTATCAACAACATGTTGAATAATGCCGCTTCGATGATA 1034295  
QY 4388 tcgcttaacaataactctgggttttcagagtaggaacttcttagctcaacaaggaactc 4447  
Dbb 1034294 TCGCTTCAATAAATCTCTGGGTTTCAGAGGTAGGAACCTTTCTTAGCTCAACAGGAACCTC 1034235

QY 4448 ctctttccgaagaattcaagtactacagccgcggaaacttcaagttgccatcgatgccaaac 4507  
DB 1034234 CTCTTTCGAAAGAAATTCAGTTACTACAGCCGCGAACTTCAGTTGCCATCGATGCCAAAC 1034175  
QY 4508 ctgacaagattttatctcctagagcgtgcatttagtaagatagtggggaaacccaagcca 4567  
DB 1034174 CTAGACAAGATTTTATCTCTAGAGCTGCAATTTAGTAAGATGGTGGGAAACCAACCAAGCCA 1034115  
QY 4568 tcaaaaaaatgcataattacttctccataaaggctcgtgagctactcttaccgaagcttctgtct 4627  
DB 1034114 TCAAAAAATGCATATTTACTTCCATAGGCGCTCTGAGTACTCTTACCAAGCTTCTGTCT 1034055  
QY 4628 atggaggtaaatctctgtatttcttctgtctcaataaacaacatggttgggcaactctcttcc 4687  
DB 1034054 ATGGAGGTAAATTCCTGTATTCTTCTGCTCAATAAGCAACATGGTTGGGCATCTCTCTTCC 1033995  
QY 4688 taatacaggagctcgtctcctatgaacatattaaacatgatacaacaacatttaccctt 4747  
DB 1033994 TAATACAAGGAGTCTGTCTCTATGACATATTAACATGATACAATACTTTTACCCTT 1033935  
QY 4748 ctatccatgaaagaataaaggagattgggaagatttaggattggttagcggaattcttctgta 4807  
DB 1033934 CTATCCATGAAGAAATAAAGAGATTGGGAAGATTTAGGATGGTTACGGATCTCTCGTA 1033875  
QY 4808 tctctatggatcttaagaacacctctaaagattcttctaaacggatcactgtctctatg99g 4867  
DB 1033874 TCTCTATGGATCTTAAAGAACCTTCTAAAGATTTCTTAAACGGATCACTGCTATGGGG 1033815  
QY 4868 aactcgattctccagcattcccgagaacagttcacagaataattacgacccaagac 4927  
DB 1033814 AACTCGAGTATTCAGCAATTCGCCAGAACAGTTACAGAAATCGATTACCATCAAGAC 1033755  
QY 4928 acttcgatgattgcttacagaaaactctgccttctcctgtggatgcgctgcgaaggag 4987  
DB 1033754 ACTTCGATGATTGCTTACAGAAATCTGCTGCTTCTGCGATGCGCTGTCGAAGGAG 1033695  
QY 4988 ctatcatgaactgttaatatctctatgtataataaagcttgcatagcctacatgccttcta 5047  
DB 1033694 CTATCATGAACCTGAATATTTCTATGTATAATAAGCTTGCAATTAGCCFACATGCCCTTCTA 1033635  
QY 5048 tctacagaataatcctctgtctaaatcaggtattgtcttcgaatgaagctgccaag 5107  
DB 1033634 TCTACAGAAATAATCTGCTGTAAATATCGGGTATTGCTTCGAATGAAGCTGGTCAAG 1033575  
QY 5108 ttatctgcggagtgcgaactagaaactctgtctagagcagaatacacagtcaactatatac 5167  
DB 1033574 TTATCTCGGGAGTSCCAACTAGAACCTCTGCTAGAGCAGATACAGTACTCAACTATATC 1033515  
QY 5168 ttgggtccctcttggaactctctacgaaactatactatcgatgtaggcattgtatagcgtat 5227  
DB 1033514 TTGGTCCCTTCTGGACTCTCTACGGAACCTATACTATCGATGAGGCATGTATAC-CTAT 1033456  
QY 5228 cgaaatgactagctgcggtgctcgcgatgtctttaa 5265  
DB 1033455 CGAAATGACTAGTGGGCTGCTGCATGATCTCTCTAA 1033418

RESULT 3  
ID AAA64755 standard; DNA; 5331 BP.  
XX  
AC AAA64755;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE C. trachomatis pmpC gene coding sequence.  
XX Chlamydial infection; sexually transmitted disease;  
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
KW trachoma; blindness; acute respiratory tract infection;  
KW atherosclerosis; coronary heart disease; antibacterial; ss.  
OS Chlamydia trachomatis.

XX WO200034483-A2.  
PN 15-JUN-2000.  
XX  
PD 08-DEC-1999; 99WO-US29012.  
XX  
PF 08-DEC-1998; 98US-0208277.  
XX  
PR 08-APR-1999; 99US-0288594.  
PR 01-OCT-1999; 99US-0410568.  
PR 22-OCT-1999; 99US-0426571.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
XX WPI; 2000-431303/37.  
XX  
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
PT comprises immunogenic portion of Chlamydia antigen, which comprises  
PT amino acid sequence encoded by polynucleotide sequence -  
XX  
PS Claim 1; Pages 176-177; 256pp; English.  
XX  
CC The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamydiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and can  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is a  
CC the leading cause of preventable blindness worldwide. C. pneumonia is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a nucleic acid sequence  
CC isolated in the present invention.  
XX  
SQ Sequence 5331 BP; 1734 A; 1170 C; 1062 G; 1365 T; 0 other;

Query Match 15.8%; Score 830.6; DB 21; Length 5331;  
Best Local Similarity 59.1%; Pred. No. 5.1e-210;  
Matches 1769; Conservative 0; Mismatches 1014; Indels 210; Gaps 12;

QY 2426 cagcttcagcactcgtcgagctgtgtcttccctacaagcagccgagcagcgacacat 2485  
DB 2396 cagactcttccctcgtggagatagcgtggagactctgaagaccgactgagccagaag 2455  
QY 2486 catctccagcaacacacaaacttattcaggtgttagtaggagagctatctatgagaaaaag 2545  
DB 2456 ctggttctacaacagaactcctactttaataggaggaggtgctatctatgagaaaactg 2515  
QY 2546 ttacattctctcaatgtagcgggacttgcgtctctctgggaaccaagctatcgataaca 2605  
DB 2516 ttaagattgagaactctctcgtgccaaggatattttctggaaacaaagctatcgataaca 2575  
QY 2606 -----atccctcccaatcatcgttgaaagctacaagagagagccatctatgccaaaaact 2659  
DB 2576 ccacagaaggctcctcttccaaatcctaaactcgtcgtggaggtcggttctatgctaaacat 2635  
QY 2660 ctttgcatttgatcttccgatgctgggaacctcctatatttctcggggaacagtgctct 2719  
DB 2636 tgtttaatctcatagcgggagcttagcgaactgtcacccttctccgggaataactgtct 2695  
QY 2720 ccactggggaatctcaacaacacaggggcaaatagcgggagagcgatctactccctactg 2779  
DB 2696 cttct-----caatctacaacaggtcaggttgctggaggagctatctactctctactg 2749  
QY 2780 ttacattgaattgtcctcggaactctctcaacaatcacagcctctatagctacacgaaga 2839  
DB 2750 taaccattgctactcctgtagttattttctaaaaaactctgcaacaaactgcttaataacg 2809







XX Chlamydia infection; sexually transmitted disease;  
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
KW trachoma; blindness; acute respiratory tract infection;  
KW atherosclerosis; coronary heart disease; antibacterial; ss.  
XX Chlamydia trachomatis.  
XX WO200034483-A2.  
XX 15-JUN-2000.  
XX 08-DEC-1999; 99WO-US29012.  
XX 08-DEC-1998; 98US-0208277.  
XX 08-APR-1999; 99US-0288594.  
XX 01-OCT-1999; 99US-0410568.  
XX 22-OCT-1999; 99US-0426571.  
XX (CORI-) CORIXA CORP.  
XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
XX WPI; 2000-431303/37.  
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
PT comprises immunogenic portion of Chlamydia antigen, which comprises  
PT amino acid sequence encoded by polynucleotide sequence -  
XX Claim 1; Page 240; 256pp; English.  
XX The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamydiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and can  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a nucleic acid sequence  
CC isolated in the present invention.  
XX Sequence 598 BP; 200 A; 126 C; 123 G; 149 T; 0 other;  
XX Query Match 10.7%; Score 564; DB 21; Length 598;  
XX Best Local Similarity 98.8%; Pred. NO. 9.5e-140;  
XX Matches 589; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
QY 1209 cctctctctcctcaatcgttcttgagcaactacagctcccgactcaggagactctag 1268  
DB 1 cctctctctcctcaatcgttcttgagcaactacagctcccgactcaggagactctag 60  
QY 1269 cctcggctcagactcggtacacgtccagaaacagttccagtcacagctaaaggcgtggtc 1328  
DB 61 cctcggctcaactcggatactcctcaaaaacagttccagtcacagctaaaggcgtggtc 120  
QY 1329 ttatactgataagaattcttcgattactaatacatcacaggaaattatcgaaattgcaataa 1388  
DB 121 ttatactgataagaattcttcgattactaatacatcacaggaaattatcgaaattgcaataa 180  
QY 1389 caaagcgacagatgttgagggtggtgcttacgtacgtacgtacgtacgtacgtacgtacgt 1448  
DB 181 caaagcgacagatgttgagggtggtgcttacgtacgtacgtacgtacgtacgtacgtacgt 240  
QY 1449 tcacgcctcacatattttgaaacactctccgataaacaagtgagggaattacagcaga 1508  
DB 241 tcacgcctcacatattttgaaacactctccgataaacaagtgagggaattacagcaga 300

QY 1509 agacaacatcacctatctaatattgacaggagactctattccaagaataactgcca 1568  
DB 301 agacaacatcacctatctaatattgacaggagactctattccaagaataactgcca 360  
QY 1569 agaagaggcggtgactcttataaaaaggtacagataaagctcttacaatgacagact 1628  
DB 361 aaagaggcggtgactcttataaaaaggtacagataaagctcttacaatgacagact 420  
QY 1629 ggatgcttctcttataataataacacatcacagaaacatggtggt-ggagcctttgtta 1687  
DB 421 ggatgcttctcttataataataacacatcacagaaacatggtggtggtggtggtttgta 480  
QY 1688 ccaagaaatctctcagacttacacctc-tgatgtggaacaaattccaggaatcacgct 1746  
DB 481 ccaagaaatctctcagacttacacctc-tgatgtggaacaaattccaggaatcacgct 540  
QY 1747 gacatggtgaaacagctcattactggaataataatcacaggaggttaattggtggagg 1802  
DB 541 gacatggtgaaacagctcattactggaataataatcacaggaggttaattggtggagg 596  
RESULT 6  
AAA64796/C  
ID AAA64796 standard; DNA; 511 BP.  
XX AC AAA64796;  
XX DT 02-FEB-2001 (first entry)  
XX DE C. trachomatis 5' end CcL2gam-30 coding sequence.  
XX KW Chlamydial infection; sexually transmitted disease;  
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
KW trachoma; blindness; acute respiratory tract infection;  
KW atherosclerosis; coronary heart disease; antibacterial; ss.  
XX OS Chlamydia trachomatis.  
XX PN WO200034483-A2.  
XX PD 15-JUN-2000.  
XX PF 08-DEC-1999; 99WO-US29012.  
XX PR 08-DEC-1998; 98US-0208277.  
XX PR 08-APR-1999; 99US-0288594.  
XX PR 01-OCT-1999; 99US-0410568.  
XX PR 22-OCT-1999; 99US-0426571.  
XX PA (CORI-) CORIXA CORP.  
XX PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
XX WPI; 2000-431303/37.  
XX DR Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
XX PT comprises immunogenic portion of Chlamydia antigen, which comprises  
XX PT amino acid sequence encoded by polynucleotide sequence -  
XX PS Claim 1; Page 240; 256pp; English.  
XX The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamydiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and can  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a nucleic acid sequence  
CC isolated in the present invention.  
XX Sequence 598 BP; 200 A; 126 C; 123 G; 149 T; 0 other;  
XX Query Match 10.7%; Score 564; DB 21; Length 598;  
XX Best Local Similarity 98.8%; Pred. NO. 9.5e-140;  
XX Matches 589; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
QY 1209 cctctctctcctcaatcgttcttgagcaactacagctcccgactcaggagactctag 1268  
DB 1 cctctctctcctcaatcgttcttgagcaactacagctcccgactcaggagactctag 60  
QY 1269 cctcggctcagactcggtacacgtccagaaacagttccagtcacagctaaaggcgtggtc 1328  
DB 61 cctcggctcaactcggatactcctcaaaaacagttccagtcacagctaaaggcgtggtc 120  
QY 1329 ttatactgataagaattcttcgattactaatacatcacaggaaattatcgaaattgcaataa 1388  
DB 121 ttatactgataagaattcttcgattactaatacatcacaggaaattatcgaaattgcaataa 180  
QY 1389 caaagcgacagatgttgagggtggtgcttacgtacgtacgtacgtacgtacgtacgtacgt 1448  
DB 181 caaagcgacagatgttgagggtggtgcttacgtacgtacgtacgtacgtacgtacgtacgt 240  
QY 1449 tcacgcctcacatattttgaaacactctccgataaacaagtgagggaattacagcaga 1508  
DB 241 tcacgcctcacatattttgaaacactctccgataaacaagtgagggaattacagcaga 300

```
CC coronary heart disease. The present sequence is a nucleic acid sequence
XX isolated in the present invention.
SQ Sequence 511 BP; 132 A; 101 C; 131 G; 145 T; 2 other;

Query Match          9.0%; Score 472.2; DB 21; Length 511;
Best Local Similarity 98.8%; Pred. No. 2.1e-115;
Matches 485; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 2216 ctgagaactcgcgtacagagataggtggaggtatctgctgtaagaatctttagaactag 2275
Db 511 CTGAGAACTCGGCTACAGAGATAGGTGGAGGTATCTGCTGTAAGAAATCTTTAAACTAG 452

Qy 2276 atgc-tctagtccttatctgaacagagacacctgtgtggaagaaggtggagccta 2334
Db 451 ATGNTCTAGTCTCTTAATCTGTACAGAGAACCTTGTGGGAAGAGGTGGAGGCTTA 392

Qy 2335 catgctaaactgtaaattttctaattctgaatcaggtctcttcttcgaaacacaaa 2394
Db 391 CATGCTAAACTGTAAATATTTCTAATCTGAATCAGGCTTCTTCTTCGAACAACAAA 332

Qy 2395 gcaactcctcatcacagaggtcgcaacaacagcttcagcacctgtgagctgctgct 2454
Db 331 GCAAACTCTCATCCACAGAGTGCACAAACAGCTTCAGCACCTGCTGCAGCTGCTGCT 272

Qy 2455 tccctaagcagcgcgcagcgcgcacccatcatctccagcaacacacacacttaccgt 2514
Db 271 TCCCTACAAAGCAGCGCAGCAGCGCACCATCATCTCCAGCAACACCAACTATTTCAGGT 212

Qy 2515 gtatgagagagctatctatggagaaaaggttacattctctcaatgtagcgggacttgt 2574
Db 211 GTAGTAGGAGGAGCTATCTATGGAGAACAGCTTACATTCTCTCAATGTAGCGGACTTGT 152

Qy 2575 cagttctctgggaacacagctatcgatatacaaatccctcccaatcatcgttgaagctacaa 2634
Db 151 CAGTTCTCTGGGAACCAACAGCTATCGATACAAATCCCTCCCAATCATCTGTTGAAGGTACAA 92

Qy 2635 ggagagcctatctatgcacaaacctttgtctattggtatctccgatctcgtggaacctcc 2694
Db 91 GGAGGAGGCATCTATGCCAAAACCTCTTTGTCTATTAGATCTTCGATGWTGGAACTCC 32

Qy 2695 tatattttctc 2705
Db 31 TATATTTTCTC 21

RESULT 7
AAA50034
ID AAA50034 standard; DNA; 1550 BP.
XX
AC AAA50034;
XX
DT 10-OCT-2000 (first entry)
XX
DE DNA encoding Chlamydia pneumoniae antigen CPN100711 RY-59.
XX
KW CPN100711 RY-59; antigen; infection; diagnosis; therapy; vaccine;
KW outer membrane protein; ds.
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 10..1419
FT /*tag= a
XX
XX WO200039158-A1.
XX
PN 23-DEC-1999; 99WO-CA01230.
PD 06-JUL-2000.
XX
XX 23-DEC-1998; 98US-0113280.
```

```
PR 23-DEC-1998; 98US-0113281.
PR 23-DEC-1998; 98US-0113282.
PR 23-DEC-1998; 98US-0113283.
PR 23-DEC-1998; 98US-0113284.
PR 23-DEC-1998; 98US-0113285.
PR 23-DEC-1998; 98US-0113385.
PR 23-DEC-1998; 98US-0113385.
PR 28-DEC-1998; 98US-0114050.
PR 28-DEC-1998; 98US-0114056.
PR 28-DEC-1998; 98US-0114057.
PR 28-DEC-1998; 98US-0114058.
PR 28-DEC-1998; 98US-0114059.
PR 28-DEC-1998; 98US-0114061.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX
DR WPI; 2000-452369/39.
DR P-PSDB; AAY95547.
XX
XX Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
XX prevention and treatment of Chlamydia infection in mammals -
XX
XX Claim 2(a); Page 62-65; 215pp; English.
XX
XX The present sequence is that of Chlamydia pneumoniae genomic DNA
XX including an open reading frame that codes for CPN100711 RY-59 (see
XX AAY95547), a putative outer membrane protein. It is an example of
XX C. pneumoniae polynucleotide molecules of the invention (see
XX AA50030-42) that encode antigenic polypeptides (see AAY95543-55) useful
XX in the diagnosis, treatment and prevention of Chlamydia infection.
XX The polynucleotides can be utilised: in the recombinant production
XX of Chlamydia antigens using transformed unicellular host cells; in
XX vaccines or live vaccine vectors; in naked form or formulated with
XX a delivery vehicle for therapy and prophylaxis of Chlamydia
XX infection; and as probes or primers for diagnosis of Chlamydia
XX infection.
XX
SQ Sequence 1550 BP; 467 A; 341 C; 306 G; 436 T; 0 other;

Query Match          8.3%; Score 435.8; DB 21; Length 1550;
Best Local Similarity 60.3%; Pred. No. 1.8e-105;
Matches 763; Conservative 0; Mismatches 487; Indels 15; Gaps 2;

Qy 4003 agcgagaacaaagtttcttaacagagagacaccttatttaatatagatcctaaggaaacttt 4062
Db 166 agtaagataagattgatattacaggaactgtgactctctctagatcctaattggcaactta 225

Qy 4063 taccaaaaccctatgttaggaagcgatctagatgtaccactaattaaagcttccgactaac 4122
Db 226 tatcaaaattcttattctgttggaagacgcgatcatcactcttttcaatatagacaattct 285

Qy 4123 acaagtgcac---gtccaagtctatgatttaactttatctctggggtcttttccctcagaaa 4179
Db 286 gcaagtggggcagttacagccacgaatgtccacctcaagggaatttaggagagctaaaaaa 345

Qy 4180 gggtacatgggaaccttggaacatttagatttctaataccacacacacacacacacacacac 4239
Db 346 ggatatttaggaacctggaaatttggtaccaattctcctcggttccaaaaattattctaaaa 405

Qy 4240 tggcatttcgataccctcgtcgctgggtatatacatcactaggtggaataatcttttttcgcg 4299
Db 406 tggacctttgacaaataactcgtcgctggcctacatccctagacacacacacttctacac 465

Qy 4300 aactctatcttagctcccaaaactcaatgatgtgtggaagcagggtctatcaacaac 4359
Db 466 aactctatttgggagacacaaactctttagtgcgtggaacacaggggtctcttaggggac 525

Qy 4360 atgttgaataatgccgcttcgatgatcgtcttacaataaactctctgggttttcaggagta 4419
Db 526 atgttgaacaatgcaagggttgaagatcctcttcaacaacactctctgggttcggtcgata 585
```

QY 4420 ggaactttcttagctcaacaagaactctcttctccgaagaattcctagttactacagccgc 4479  
Db 586 ggaactttcttaggaagaagatctctgaattctgactcattccattcagtcagcaga 645  
QY 4480 ggaacttcagttgcatcagaccacacagatagacaagattttatctcagagctgatit 4539  
Db 646 ggcataccgcgcgtggtgagcccaacctcgcacagaattttatttttagagctgcttc 705  
QY 4540 agtaagatagtggggaaacacacacatccataaaaaatgcataattacttccataaggcc 4599  
Db 706 agtcaggtttttggtcacgcagctgctgaatcaccttgacacataagaacataaggc 765  
QY 4600 tctgagtaactcttaccagaactctctgctatgaggttaattctctgattcttctgctcaat 4659  
Db 766 tcaggctactctacacaagcatcttttatgctggaatatcttcttatcttctctcgatgata 825  
QY 4660 aagcaacatggttgccactctcttctcaataacagaagagtgctgctctatggacataatt 4719  
Db 826 c-----ggctcggtctctctcttcccaagggtgctgacacctatggttatg 873  
QY 4720 aaacatgatacaacaacactttacccttctctccatgaagaataaaagagattggaa 4779  
Db 874 caacatgacacacacacactactctctctctattgaagaaaaaataatgcacaactggat 933  
QY 4780 gatttagatggttagcgatcttctctctctctctctctctctctctctctctctctctct 4839  
Db 934 agcattgctggttattttagctgcttctctctctctctctctctctctctctctctctctct 993  
QY 4840 tctctaaacggtacactgct 4899  
Db 994 tctacagaaggtctacact 1053  
QY 4900 ttcacagaactcgattacgataccacacactctctctctctctctctctctctctctctct 4959  
Db 1054 ttcacagaactcgattacgataccacacactctctctctctctctctctctctctctctct 1113  
QY 4960 cttctctggtggtgctctgcaagagatctctctctctctctctctctctctctctctctct 5019  
Db 1114 attcctactggtattct 1173  
QY 5020 aagcttgcatagctacatgct 5079  
Db 1174 aaagtatacgtcgtacct 1233  
QY 5080 gtattgtctgaatgaagctggtcaagtattctctctctctctctctctctctctctctctct 5139  
Db 1234 gttctctctacaaaagaaaggcaacgtagtcacagttctctctctctctctctctctctct 1293  
QY 5140 agacagataacagtactcaactatctctctctctctctctctctctctctctctctctctct 5199  
Db 1294 cgtcagaggtgagctctcaatttctctctctctctctctctctctctctctctctctctct 1353  
QY 5200 actatcgatgtagcatgtatcgcctatcgcctctctctctctctctctctctctctctctct 5259  
Db 1354 acattgagctctcaatgaatacttctctctctctctctctctctctctctctctctctctct 1413  
QY 5260 ttcta 5264  
Db 1414 ttcta 1418

RESULT 8

AAx91990  
ID AAX91990 standard; DNA; 1230025 BP.  
XX  
AC AAX91990;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; ss.  
XX Chlamydia pneumoniae.  
XX WO9927105-A2.  
XX 03-JUN-1999.  
XX 20-NOV-1998; 98WO-1B01890.  
XX 04-NOV-1998; 98US-0107078.  
XX 21-NOV-1997; 97FR-0014673.  
XX (GEST ) GENSET.  
XX Griffais R;  
XX WPI; 1999-357842/30.  
XX Genome sequence of Chlamydia pneumoniae  
XX Claim 1; Page 291-611; 1912pp; English.  
XX The present sequence represents the complete genome of Chlamydia pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.  
XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 8.3%; Score 435.8; DB 20; Length 1230025;  
Best Local Similarity 60.3%; Pred. No. 7.8e-104;  
Matches 763; Conservative 0; Mismatches 487; Indels 15; Gaps 2;

QY 4003 agcgagagacaaagtttctctacagagagacaccttactttaatagatcctaatggaaacttt 4062  
Db 635707 agtaagataaagattgatattacaggaactgtgactctctctagatcctaatggcaactta 635766  
QY 4063 taccacaaaccttatgttagaagcgatctagatgaccactaattagcttcgactaac 4122  
Db 635767 tatcaaaattcttatcttggtagaacccgcgatacactcttttcaatagacaaattct 635826  
QY 4123 acaagtgac---gtccaaagtctatgatttaactttatctgggactcttttccctcagaaa 4179  
Db 635827 gcaagtggggcagttacagccacgaatgtcacaccttcaagggaatttagagctaaaaa 635886  
QY 4180 gggtaactgggaacctggacatttagattctaataccacaaacaggggaaacttcaagcaga 4239  
Db 635887 ggatatttagaaccttggatttggatccaaattctcgggttcaaaaattattctaaaa 635946  
QY 4240 tggacattcgatacctatcgctcggtggtatcacatccttaggagataatcttttatcg 4299  
Db 635947 tggacctttgacaaaatacctcgctggtccctacatccctcagagacacacctctcacatc 636006  
QY 4300 aactctatcttaggtcccaaaactcaatgattgttgaaagcaaggccttatcaacaac 4359  
Db 636007 aactctatttgggagacacaaaactctttagtgactgtgaaacagggatcttagggaac 636066  
QY 4360 atgttgaataatgcccgcttcgatgatactgcttacaataaacttctgggttttcaggaajta 4419  
Db 636067 atgttgaacaatgcaaggtttgaagatcgtctttcaacaacttctgggttcggtgata 636126  
QY 4420 ggaactttttagtcaacaagaactctcttcttcgaagaattcagttactacagcgc 4479  
Db 636127 ggaactttttaggaaagaagtatctcgaataattctgactcattccactatcatggtcgta 636186













**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2001, 18:19:12 ; Search time 3705.76 Seconds  
(without alignments)  
13430.247 Million cell updates/sec

Title: US-09-454-684A-174  
Perfect score: 5265  
Sequence: 1 gcaatcatgaatggctgtc.....gtgctcgcatgatcttcttaa 5265

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
259: gb_est190:*
260: gb_est191:*
261: gb_est192:*
262: gb_est193:*
263: gb_est194:*
264: gb_est195:*
265: gb_est196:*
266: gb_est197:*
267: gb_est198:*
268: gb_est199:*
269: gb_est200:*
270: gb_est201:*
271: gb_est202:*
272: gb_est203:*
273: gb_est204:*
274: gb_est205:*
275: gb_est206:*
276: gb_est207:*
277: gb_est208:*
278: gb_est209:*
279: gb_est210:*
280: gb_est211:*
281: gb_est212:*
282: gb_est213:*
283: gb_est214:*
284: gb_est215:*
285: gb_est216:*
286: gb_est217:*
287: gb_est218:*
288: gb_est219:*
289: gb_est220:*
290: gb_est221:*
291: gb_est222:*
292: gb_est223:*
293: gb_est224:*
294: gb_est225:*
295: gb_est226:*
296: gb_est227:*
297: gb_est228:*
298: gb_est229:*
299: gb_est230:*
300: gb_est231:*
301: gb_est232:*
302: gb_est233:*
303: gb_est234:*
304: gb_est235:*
305: gb_est236:*
306: gb_est237:*
307: gb_est238:*
308: gb_est239:*
309: gb_est240:*
310: gb_est241:*
311: gb_est242:*
312: gb_est243:*
313: gb_est244:*
314: gb_est245:*
315: gb_est246:*
316: gb_est247:*
317: gb_est248:*
318: gb_est249:*
319: gb_est250:*
320: gb_est251:*
321: gb_est252:*
322: gb_est253:*
323: gb_est254:*
324: gb_est255:*
325: gb_est256:*
326: gb_est257:*
327: gb_est258:*
328: gb_est259:*
329: gb_est260:*
330: gb_est261:*
331: gb_est262:*
332: gb_est263:*
333: gb_est264:*
334: gb_est265:*
335: gb_est266:*
336: gb_est267:*
337: gb_est268:*
338: gb_est269:*
339: gb_est270:*
340: gb_est271:*
341: gb_est272:*
342: gb_est273:*
343: gb_est274:*
344: gb_est275:*
345: gb_est276:*
346: gb_est277:*
347: gb_est278:*
348: gb_est279:*
349: gb_est280:*
350: gb_est281:*
351: gb_est282:*
352: gb_est283:*
353: gb_est284:*
354: gb_est285:*
355: gb_est286:*
356: gb_est287:*
357: gb_est288:*
358: gb_est289:*
359: gb_est290:*
360: gb_est291:*
361: gb_est292:*
362: gb_est293:*
363: gb_est294:*
364: gb_est295:*
365: gb_est296:*
366: gb_est297:*
367: gb_est298:*
368: gb_est299:*
369: gb_est300:*
370: gb_est301:*
371: gb_est302:*
372: gb_est303:*
373: gb_est304:*
374: gb_est305:*
375: gb_est306:*
376: gb_est307:*
377: gb_est308:*
378: gb_est309:*
379: gb_est310:*
380: gb_est311:*
381: gb_est312:*
382: gb_est313:*
383: gb_est314:*
384: gb_est315:*
385: gb_est316:*
386: gb_est317:*
387: gb_est318:*
388: gb_est319:*
389: gb_est320:*
390: gb_est321:*
391: gb_est322:*
392: gb_est323:*
393: gb_est324:*
394: gb_est325:*
395: gb_est326:*
396: gb_est327:*
397: gb_est328:*
398: gb_est329:*
399: gb_est330:*
400: gb_est331:*
401: gb_est332:*
402: gb_est333:*
403: gb_est334:*
404: gb_est335:*
405: gb_est336:*
406: gb_est337:*
407: gb_est338:*
408: gb_est339:*
409: gb_est340:*
410: gb_est341:*
411: gb_est342:*
412: gb_est343:*
413: gb_est344:*
414: gb_est345:*
415: gb_est346:*
416: gb_est347:*
417: gb_est348:*
418: gb_est349:*
419: gb_est350:*
420: gb_est351:*
421: gb_est352:*
422: gb_est353:*
423: gb_est354:*
424: gb_est355:*
425: gb_est356:*
426: gb_est357:*
427: gb_est358:*
428: gb_est359:*
429: gb_est360:*
430: gb_est361:*
431: gb_est362:*
432: gb_est363:*
433: gb_est364:*
434: gb_est365:*
435: gb_est366:*
436: gb_est367:*
437: gb_est368:*
438: gb_est369:*
439: gb_est370:*
440: gb_est371:*
441: gb_est372:*
442: gb_est373:*
443: gb_est374:*
444: gb_est375:*
445: gb_est376:*
446: gb_est377:*
447: gb_est378:*
448: gb_est379:*
449: gb_est380:*
450: gb_est381:*
451: gb_est382:*
452: gb_est383:*
453: gb_est384:*
454: gb_est385:*
455: gb_est386:*
456: gb_est387:*
457: gb_est388:*
458: gb_est389:*
459: gb_est390:*
460: gb_est391:*
461: gb_est392:*
462: gb_est393:*
463: gb_est394:*
464: gb_est395:*
465: gb_est396:*
466: gb_est397:*
467: gb_est398:*
468: gb_est399:*
469: gb_est400:*
470: gb_est401:*
471: gb_est402:*
472: gb_est403:*
473: gb_est404:*
474: gb_est405:*
475: gb_est406:*
476: gb_est407:*
477: gb_est408:*
478: gb_est409:*
479: gb_est410:*
480: gb_est411:*
481: gb_est412:*
482: gb_est413:*
483: gb_est414:*
484: gb_est415:*
485: gb_est416:*
486: gb_est417:*
487: gb_est418:*
488: gb_est419:*
489: gb_est420:*
490: gb_est421:*
491: gb_est422:*
492: gb_est423:*
493: gb_est424:*
494: gb_est425:*
495: gb_est426:*
496: gb_est427:*
497: gb_est428:*
498: gb_est429:*
499: gb_est430:*
500: gb_est431:*
501: gb_est432:*
502: gb_est433:*
503: gb_est434:*
504: gb_est435:*
505: gb_est436:*
506: gb_est437:*
507: gb_est438:*
508: gb_est439:*
509: gb_est440:*
510: gb_est441:*
511: gb_est442:*
512: gb_est443:*
513: gb_est444:*
514: gb_est445:*
515: gb_est446:*
516: gb_est447:*
517: gb_est448:*
518: gb_est449:*
519: gb_est450:*
520: gb_est451:*
521: gb_est452:*
522: gb_est453:*
523: gb_est454:*
524: gb_est455:*
525: gb_est456:*
526: gb_est457:*
527: gb_est458:*
528: gb_est459:*
529: gb_est460:*
530: gb_est461:*
531: gb_est462:*
532: gb_est463:*
533: gb_est464:*
534: gb_est465:*
535: gb_est466:*
536: gb_est467:*
537: gb_est468:*
538: gb_est469:*
539: gb_est470:*
540: gb_est471:*
541: gb_est472:*
542: gb_est473:*
543: gb_est474:*
544: gb_est475:*
545: gb_est476:*
546: gb_est477:*
547: gb_est478:*
548: gb_est479:*
549: gb_est480:*
550: gb_est481:*
551: gb_est482:*
552: gb_est483:*
553: gb_est484:*
554: gb_est485:*
555: gb_est486:*
556: gb_est487:*
557: gb_est488:*
558: gb_est489:*
559: gb_est490:*
560: gb_est491:*
561: gb_est492:*
562: gb_est493:*
563: gb_est494:*
564: gb_est495:*
565: gb_est496:*
566: gb_est497:*
567: gb_est498:*
568: gb_est499:*
569: gb_est500:*
570: gb_est501:*
571: gb_est502:*
572: gb_est503:*
573: gb_est504:*
574: gb_est505:*
575: gb_est506:*
576: gb_est507:*
577: gb_est508:*
578: gb_est509:*
579: gb_est510:*
580: gb_est511:*
581: gb_est512:*
582: gb_est513:*
583: gb_est514:*
584: gb_est515:*
585: gb_est516:*
586: gb_est517:*
587: gb_est518:*
588: gb_est519:*
589: gb_est520:*
590: gb_est521:*
591: gb_est522:*
592: gb_est523:*
593: gb_est524:*
594: gb_est525:*
595: gb_est526:*
596: gb_est527:*
597: gb_est528:*
598: gb_est529:*
599: gb_est530:*
600: gb_est531:*
601: gb_est532:*
602: gb_est533:*
603: gb_est534:*
604: gb_est535:*
605: gb_est536:*
606: gb_est537:*
607: gb_est538:*
608: gb_est539:*
609: gb_est540:*
610: gb_est541:*
611: gb_est542:*
612: gb_est543:*
613: gb_est544:*
614: gb_est545:*
615: gb_est546:*
616: gb_est547:*
617: gb_est548:*
618: gb_est549:*
619: gb_est550:*
620: gb_est551:*
621: gb_est552:*
622: gb_est553:*
623: gb_est554:*
624: gb_est555:*
625: gb_est556:*
626: gb_est557:*
627: gb_est558:*
628: gb_est559:*
629: gb_est560:*
630: gb_est561:*
631: gb_est562:*
632: gb_est563:*
633: gb_est564:*
634: gb_est565:*
635: gb_est566:*
636: gb_est567:*
637: gb_est568:*
638: gb_est569:*
639: gb_est570:*
640: gb_est571:*
641: gb_est572:*
642: gb_est573:*
643: gb_est574:*
644: gb_est575:*
645: gb_est576:*
646: gb_est577:*
647: gb_est578:*
648: gb_est579:*
649: gb_est580:*
650: gb_est581:*
651: gb_est582:*
652: gb_est583:*
653: gb_est584:*
654: gb_est585:*
655: gb_est586:*
656: gb_est587:*
657: gb_est588:*
658: gb_est589:*
659: gb_est590:*
660: gb_est591:*
661: gb_est592:*
662: gb_est593:*
663: gb_est594:*
664: gb_est595:*
665: gb_est596:*
666: gb_est597:*
667: gb_est598:*
668: gb_est599:*
669: gb_est600:*
670: gb_est601:*
671: gb_est602:*
672: gb_est603:*
673: gb_est604:*
674: gb_est605:*
675: gb_est606:*
676: gb_est607:*
677: gb_est608:*
678: gb_est609:*
679: gb_est610:*
680: gb_est611:*
681: gb_est612:*
682: gb_est613:*
683: gb_est614:*
684: gb_est615:*
685: gb_est616:*
686: gb_est617:*
687: gb_est618:*
688: gb_est619:*
689: gb_est620:*
690: gb_est621:*
691: gb_est622:*
692: gb_est623:*
693: gb_est624:*
694: gb_est625:*
695: gb_est626:*
696: gb_est627:*
697: gb_est628:*
698: gb_est629:*
699: gb_est630:*
700: gb_est631:*
701: gb_est632:*
702: gb_est633:*
703: gb_est634:*
704: gb_est635:*
705: gb_est636:*
706: gb_est637:*
707: gb_est638:*
708: gb_est639:*
709: gb_est640:*
710: gb_est641:*
711: gb_est642:*
712: gb_est643:*
713: gb_est644:*
714: gb_est645:*
715: gb_est646:*
716: gb_est647:*
717: gb_est648:*
718: gb_est649:*
719: gb_est650:*
720: gb_est651:*
721: gb_est652:*
722: gb_est653:*
723: gb_est654:*
724: gb_est655:*
725: gb_est656:*
726: gb_est657:*
727: gb_est658:*
728: gb_est659:*
729: gb_est660:*
730: gb_est661:*
731: gb_est662:*
732: gb_est663:*
733: gb_est664:*
734: gb_est665:*
735: gb_est666:*
736: gb_est667:*
737: gb_est668:*
738: gb_est669:*
739: gb_est670:*
740: gb_est671:*
741: gb_est672:*
742: gb_est673:*
743: gb_est674:*
744: gb_est675:*
745: gb_est676:*
746: gb_est677:*
747: gb_est678:*
748: gb_est679:*
749: gb_est680:*
750: gb_est681:*
751: gb_est682:*
752: gb_est683:*
753: gb_est684:*
754: gb_est685:*
755: gb_est686:*
756: gb_est687:*
757: gb_est688:*
758: gb_est689:*
759: gb_est690:*
760: gb_est691:*
761: gb_est692:*
762: gb_est693:*
763: gb_est694:*
764: gb_est695:*
765: gb_est696:*
766: gb_est697:*
767: gb_est698:*
768: gb_est699:*
769: gb_est700:*
770: gb_est701:*
771: gb_est702:*
772: gb_est703:*
773: gb_est704:*
774: gb_est705:*
775: gb_est706:*
776: gb_est707:*
777: gb_est708:*
778: gb_est709:*
779: gb_est710:*
780: gb_est711:*
781: gb_est712:*
782: gb_est713:*
783: gb_est714:*
784: gb_est715:*
785: gb_est716:*
786: gb_est717:*
787: gb_est718:*
788: gb_est719:*
789: gb_est720:*
790: gb_est721:*
791: gb_est722:*
792: gb_est723:*
793: gb_est724:*
794: gb_est725:*
795: gb_est726:*
796: gb_est727:*
797: gb_est728:*
798: gb_est729:*
799: gb_est730:*
800: gb_est731:*
801: gb_est732:*
802: gb_est733:*
803: gb_est734:*
804: gb_est735:*
805: gb_est736:*
806: gb_est737:*
807: gb_est738:*
808: gb_est739:*
809: gb_est740:*
810: gb_est741:*
811: gb_est742:*
812: gb_est743:*
813: gb_est744:*
814: gb_est745:*
815: gb_est746:*
816: gb_est747:*
817: gb_est748:*
818: gb_est749:*
819: gb_est750:*
820: gb_est751:*
821: gb_est752:*
822: gb_est753:*
823: gb_est754:*
824: gb_est755:*
825: gb_est756:*
826: gb_est757:*
827: gb_est758:*
828: gb_est759:*
829: gb_est760:*
830: gb_est761:*
831: gb_est762:*
832: gb_est763:*
833: gb_est764:*
834: gb_est765:*
835: gb_est766:*
836: gb_est767:*
837: gb_est768:*
838: gb_est769:*
839: gb_est770:*
840: gb_est771:*
841: gb_est772:*
842: gb_est773:*
843: gb_est774:*
844: gb_est775:*
845: gb_est776:*
846: gb_est777:*
847: gb_est778:*
848: gb_est779:*
849: gb_est780:*
850: gb_est781:*
851: gb_est782:*
852: gb_est783:*
853: gb_est784:*
854: gb_est785:*
855: gb_est786:*
856: gb_est787:*
857: gb_est788:*
858: gb_est789:*
859: gb_est790:*
860: gb_est791:*
861: gb_est792:*
862: gb_est793:*
863: gb_est794:*
864: gb_est795:*
865: gb_est796:*
866: gb_est797:*
867: gb_est798:*
868: gb_est799:*
869: gb_est800:*
870: gb_est801:*
871: gb_est802:*
872: gb_est803:*
873: gb_est804:*
874: gb_est805:*
875: gb_est806:*
876: gb_est807:*
877: gb_est808:*
878: gb_est809:*
879: gb_est810:*
880: gb_est811:*
881: gb_est812:*
882: gb_est813:*
883: gb_est814:*
884: gb_est815:*
885: gb_est816:*
886: gb_est817:*
887: gb_est818:*
888: gb_est819:*
889: gb_est820:*
890: gb_est821:*
891: gb_est822:*
892: gb_est823:*
893: gb_est824:*
894: gb_est825:*
895: gb_est826:*
896: gb_est827:*
897: gb_est828:*
898: gb_est829:*
899: gb_est830:*
900: gb_est831:*
901: gb_est832:*
902: gb_est833:*
903: gb_est834:*
904: gb_est835:*
905: gb_est836:*
906: gb_est837:*
907: gb_est838:*
908: gb_est839:*
909: gb_est840:*
910: gb_est841:*
911: gb_est842:*
912: gb_est843:*
913: gb_est844:*
914: gb_est845:*
915: gb_est846:*
916: gb_est847:*
917: gb_est848:*
918: gb_est849:*
919: gb_est850:*
920: gb_est851:*
921: gb_est852:*
922: gb_est853:*
923: gb_est854:*
924: gb_est855:*
925: gb_est856:*
926: gb_est857:*
927: gb_est858:*
928: gb_est859:*
929: gb_est860:*
930: gb_est861:*
931: gb_est862:*
932: gb_est863:*
933: gb_est864:*
934: gb_est865:*
935: gb_est866:*
936: gb_est867:*
937: gb_est868:*
938: gb_est869:*
939: gb_est870:*
940: gb_est871:*
941: gb_est872:*
942: gb_est873:*
943: gb_est874:*
944: gb_est875:*
945: gb_est876:*
946: gb_est877:*
947: gb_est878:*
948: gb_est879:*
949: gb_est880:*
950: gb_est881:*
951: gb_est882:*
952: gb_est883:*
953: gb_est884:*
954: gb_est885:*
955: gb_est886:*
956: gb_est887:*
957: gb_est888:*
958: gb_est889:*
959: gb_est890:*
960: gb_est891:*
961: gb_est892:*
962: gb_est893:*
963: gb_est894:*
964: gb_est895:*
965: gb_est896:*
966: gb_est897:*
967: gb_est898:*
968: gb_est899:*
969: gb_est900:*
970: gb_est901:*
971: gb_est902:*
972: gb_est903:*
973: gb_est904:*
974: gb_est905:*
975: gb_est906:*
976: gb_est907:*
977: gb_est908:*
978: gb_est909:*
979: gb_est910:*
980: gb_est911:*
981: gb_est912:*
982: gb_est913:*
983: gb_est914:*
984: gb_est915:*
985: gb_est916:*
986: gb_est917:*
987: gb_est918:*
988: gb_est919:*
989: gb_est920:*
990: gb_est921:*
991: gb_est922:*
992: gb_est923:*
993: gb_est924:*
994: gb_est925:*
995: gb_est926:*
996: gb_est927:*
997: gb_est928:*
998: gb_est929:*
999: gb_est930:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83.4	1.6	252	218	AF087168	AF087168 AF087168
2	63.6	1.2	946	15	AI069309	AI069309 mgae0006c
3	61.2	1.2	978	221	CNS040R1	AL269254 Tetraodon
4	61	1.2	885	152	BG368981	BG368981 HVSMEI002
5	61	1.2	1003	221	CNS040U2	AL303092 Tetraodon
6	60.6	1.2	536	222	FR0036552	AL124061 Fugu rubr
7	60.2	1.1	906	220	CNS02ILL	AL199074 Tetraodon
8	60.2	1.1	962	222	CNS054Y6	AL321351 Tetraodon
9	58.6	1.1	808	220	CNS02NGR	AL205380 Tetraodon
10	58.4	1.1	269	247	AZ662861	AZ662861 IM0542M10
11	57.8	1.1	615	247	AZ661546	AZ661546 IM0540M01
12	57.8	1.1	712	247	AZ661544	AZ661544 IM0540M01
13	57.8	1.1	2199	139	BE741896	BE741896 601595356
14	57.6	1.1	770	139	BE744403	BE744403 601576651
15	57.2	1.1	667	221	CNS03YK	AL266285 Tetraodon
16	57	1.1	415	250	AZ839702	AZ839702 2M0135N24
17	56.6	1.1	1587	152	BG338956	BG338956 602436760
18	56	1.1	987	132	AK013560	AK013560 Mus muscu
19	55.8	1.1	727	231	AZ870651	AZ870651 2M0183011
20	55.6	1.1	562	240	AZ251135	AZ251135 RPI-23-5
21	55.6	1.1	700	105	AL508461	AL508461 AL508461
22	55.4	1.1	405	1	AA065249	AA065249 al2501r T
23	55.4	1.1	439	188	R98242	R98242 yq7se03.rl
24	55.2	1.0	455	155	BG561862	BG561862 EcEtree03
25	55	1.0	226	250	AZ837781	AZ837781 2M0133011
26	54.8	1.0	489	146	BF287438	BF287438 EST452029
27	54.8	1.0	925	140	BE778075	BE778075 601463186
28	54.6	1.0	966	219	CNS004W2	AL055504 Drosophil
29	54.6	1.0	980	219	CNS01PHD	AL154817 Anopheles
30	54.4	1.0	264	140	BE798211	BE798211 601585978
31	54.4	1.0	351	222	FR0004775	288558 F.rubripes
32	54.4	1.0	555	241	AZ281461	AZ281461 RPI-23-1
33	54.4	1.0	619	222	FR0013713	AL004959 F.rubripe
34	54.4	1.0	993	220	CNS03AOO	AL235545 Tetraodon
35	54.2	1.0	599	163	BE115577	BE115577 UR-R-BJ1-
36	54	1.0	540	116	AW464269	AW464269 BF230015A
37	54	1.0	648	18	AI257238	AI257238 LP05546.5
38	54	1.0	987	122	AW940443	AW940443 GH12043.3
39	53.8	1.0	250	115	AW409026	AW409026 fbl_d10 F
40	53.8	1.0	935	220	CNS033D4	AL225985 Tetraodon
41	53.6	1.0	162	222	FR0009521	AL000805 F.rubripe
42	53.4	1.0	354	116	AW486593	AW486593 75069 MAR
43	53.4	1.0	477	107	AU084733	AU084733 AU084733
44	53.4	1.0	680	19	AI388163	AI388163 GH18971.5
45	53.4	1.0	691	246	AZ627417	AZ627417 IM0469P07

## ALIGNMENTS

RESULT 1	AF087168	AF087168	252 bp	DNA	GSS	18-APR-2000
LOCUS	AF087168	Chlamydia trachomatis	L2	Chlamydia trachomatis	genomic	
DEFINITION	AF087168	clone 205 similar to type 1-related protein phosphatase, DNA				
ACCESSION	AF087168					
VERSION	AF087168.1	GI:7593836				
KEYWORDS	GSS					
SOURCE	Chlamydia trachomatis					
ORGANISM	Chlamydia trachomatis					
REFERENCE	1	(bases 1 to 252)				
AUTHORS	Wang, L., Steenburg, S.D., Zheng, Y. and Larsen, S.H.					
TITLE	Gene identification of Chlamydia trachomatis by random DNA					
JOURNAL	Unpublished (1998)					

## COMMENT

Contact: Wang L  
Department of Microbiology & Immunology  
Indiana University School of Medicine  
635 Barnhill Drive; MS 255, Indianapolis, IN 46202, USA  
Class: shotgun.

## FEATURES

source  
Location/Qualifiers  
1..252  
/organism="Chlamydia trachomatis"  
/strain="L2"  
/db\_xref="taxon:813"  
/clone="205"  
/clone\_lib="Chlamydia trachomatis L2"  
/note="isolate=434B"  
77 a 54 c 54 g 67 t  
BASE COUNT  
ORIGIN

Query Match 1.6%; Score 83.4; DB 218; Length 252;  
Best Local Similarity 87.9%; Pred. No. 3e-11;  
Matches 138; Conservative 0; Mismatches 11; Indels 8; Gaps 4;  
QY 4460 aattcagttactacagccggaacttcagttgcccacatg----ccaaacctagacaa 4515  
|||||  
Db 93 AATTCAGTTACTACACCCGCGAATCTCAAGTTGCCATCAGGATGCCAAACCTAGACAA 152  
|||||  
QY 4516 ga-ttttattcctagagagctgatttagtaagtagtggggaacacaaagccatcaaaa 4574  
|||||  
Db 153 GATTTTATCCTAGGAGCTGCATTTAGTAGATGGTGGGAAACCAACCAAGCCATCAANA 212  
|||||  
QY 4575 aatgcataattacttcataaggctctgagtactct 4611  
|||||  
Db 213 AATGCAT-ATTACTTCCATA--GGCTCTGAGTACTCT 246  
|||||

## RESULT 2

AI069309 946 bp mRNA EST 09-DEC-1999  
LOCUS mgae0006cc09f Magnaporthe grisea Appressorium Stage cDNA Library  
DEFINITION Magnaporthe grisea cDNA clone mgae0006cc09f 5', mRNA sequence.  
ACCESSION AI069309  
VERSION AI069309.1 GI:3392284  
KEYWORDS EST.  
SOURCE Magnaporthe grisea.  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.  
REFERENCE 1 (bases 1 to 946)  
AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.  
TITLE Expressed sequence characterization during appressorium formation in rice blast fungus, Magnaporthe grisea  
JOURNAL Unpublished (1998)  
COMMENT Contact: Dean, R.A.  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 5737  
Fax: 864 656 4293  
Email: rdean@clemson.edu  
Seq primer: r3 primer (AATTAAACCTCACTAAAGGG)  
High quality sequence stop: 403.  
Location/Qualifiers  
1..946  
/organism="Magnaporthe grisea"  
/strain="70-15"  
/db\_xref="taxon:148305"  
/clone="mgae0006cc09f"  
/clone\_lib="Magnaporthe grisea Appressorium Stage cDNA Library"  
/dev\_stage="Germinated conidia on appressorium-inductive surface"  
/note="Vector: pBlueScriptII SK(+) Vector; Site1: EcoRI; Site2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8

## FEATURES

source  
Location/Qualifiers  
1..946  
/organism="Magnaporthe grisea"  
/strain="70-15"  
/db\_xref="taxon:148305"  
/clone="mgae0006cc09f"  
/clone\_lib="Magnaporthe grisea Appressorium Stage cDNA Library"  
/dev\_stage="Germinated conidia on appressorium-inductive surface"  
/note="Vector: pBlueScriptII SK(+) Vector; Site1: EcoRI; Site2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8















Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	61.2	1.2	397	3	US-09-253-691-3	Sequence 3, Appli
2	58.6	1.1	203	4	US-09-043-303-7	Sequence 7, Appli
3	58	1.1	543	6	5273901-6	Patent No. 5273901
4	57.2	1.1	195	1	US-08-469-802B-2	Sequence 2, Appli
5	57.2	1.1	195	2	US-08-267-803B-2	Sequence 2, Appli
6	57.2	1.1	234	1	US-08-469-802B-3	Sequence 3, Appli
7	57.2	1.1	234	2	US-08-267-803B-3	Sequence 3, Appli
8	56.6	1.1	154	1	US-08-469-802B-6	Sequence 6, Appli
9	56.6	1.1	154	2	US-08-267-803B-6	Sequence 6, Appli
10	56.6	1.1	165	4	US-09-043-303-17	Sequence 17, Appli
11	56.6	1.1	168	1	US-08-469-802B-4	Sequence 4, Appli
12	56.6	1.1	168	2	US-08-267-803B-4	Sequence 4, Appli
13	56.6	1.1	171	1	US-08-469-802B-5	Sequence 5, Appli
14	56.6	1.1	171	2	US-08-267-803B-5	Sequence 5, Appli
15	52.6	1.0	3376	1	US-08-320-559-29	Sequence 29, Appli
16	52.6	1.0	3376	3	US-08-545-860D-29	Sequence 29, Appli
17	52.6	1.0	3376	5	PCR-US94-04496-29	Sequence 29, Appli
18	49.4	0.9	506	1	US-08-469-802B-7	Sequence 7, Appli
19	49.4	0.9	506	2	US-08-267-803B-7	Sequence 7, Appli
20	49	0.9	3366	1	US-08-469-802B-1	Sequence 1, Appli
21	49	0.9	3366	2	US-08-267-803B-1	Sequence 1, Appli
22	49	0.9	10660	2	US-08-267-803B-8	Sequence 8, Appli
23	49	0.9	10660	3	US-09-041-886-16	Sequence 16, Appli
24	48.2	0.9	533	6	5482709-5	Patent No. 5482709
25	48.2	0.9	3354	3	US-08-965-903B-1	Sequence 1, Appli
26	47.8	0.9	377	1	US-08-332-766A-1	Sequence 1, Appli
27	47.8	0.9	377	2	US-08-145-617-5	Sequence 5, Appli



```

RESULT 6
US-08-469-802B-3
: Sequence 3, Application US/08469802B
: Patent No. 5741645
: GENERAL INFORMATION:
: APPLICANT: Orr, Harry T.
: APPLICANT: Ranum, Laura P.W.
: APPLICANT: Chung, Ming-yi
: APPLICANT: Zoghbi, Huda Y.
: TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
: Patent No. 5741645

```



```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-267-803B-6

Query Match 1.1%; Score 56.6; DB 2; Length 154;
Best Local Similarity 63.7%; Pred. No. 4e-06;
Matches 86; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2394 agcaaaactctctacacaggagtcgcaacaacagcttcagcacctgtcgtcagctgtgc 2453
Db 3 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 62

QY 2454 ttccctacaagcagcagcagcagccacccatctctccagcaacaccacattattcagg 2513
Db 63 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 122

QY 2514 tgtagtaggaggagc 2528
Db 123 AGCAGCAGCAGCAGCAGC 137

RESULT 10
US-09-043-303-17
; Sequence 17, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Probe
US-09-043-303-17

Query Match 1.1%; Score 56.6; DB 4; Length 165;
Best Local Similarity 63.7%; Pred. No. 4.2e-06;
Matches 86; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2394 agcaaaactctctacacaggagtcgcaacaacagcttcagcacctgtcgtcagctgtgc 2453
Db 2 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 61

QY 2454 ttccctacaagcagcagcagcagccacccatctctccagcaacaccacattattcagg 2513
Db 62 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 121

QY 2514 tgtagtaggaggagc 2528
Db 122 agcagcagcagcagc 136

RESULT 11
```

```
US-08-469-802B-4
; Sequence 4, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muetting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-4

Query Match 1.1%; Score 56.6; DB 1; Length 168;
Best Local Similarity 63.7%; Pred. No. 4.2e-06;
Matches 86; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2394 agcaaaactctctacacaggagtcgcaacaacagcttcagcacctgtcgtcagctgtgc 2453
Db 2 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 61

QY 2454 ttccctacaagcagcagcagcagccacccatctctccagcaacaccacattattcagg 2513
Db 62 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 121

QY 2514 tgtagtaggaggagc 2528
Db 122 AGCAGCAGCAGCAGCAGC 136

RESULT 12
US-08-267-803B-4
; Sequence 4, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 08:34:51 ; Search time 30.88 Seconds

(without alignments)  
3439.348 Million cell updates/sec

Title: US-09-454-684A-180

Perfect score: 8942  
Sequence: 1 MKMLSATVFAVPAVPSVSGF.....IDVGMVTLQMTSCGARMIF 1752

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0601:\*

1:	/SIDSR/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2:	/SIDSR/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3:	/SIDSR/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4:	/SIDSR/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5:	/SIDSR/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6:	/SIDSR/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7:	/SIDSR/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8:	/SIDSR/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9:	/SIDSR/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10:	/SIDSR/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11:	/SIDSR/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12:	/SIDSR/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13:	/SIDSR/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14:	/SIDSR/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15:	/SIDSR/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16:	/SIDSR/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17:	/SIDSR/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18:	/SIDSR/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19:	/SIDSR/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20:	/SIDSR/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21:	/SIDSR/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22:	/SIDSR/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8942	100.0	1752	21	AA13637
2	5858	65.5	1194	20	AA137825
3	3592	40.2	1776	21	AA13636
4	2703	30.2	948	21	AA13643
5	2621	29.3	892	20	AA137822
6	2568.5	28.7	568	20	AA137828
7	2474.5	27.7	1146	20	AA135162
8	1272.5	14.2	469	21	AA195547
9	1002	11.2	915	20	AA137823
10	943.5	10.6	821	21	AA13644
11	748.5	8.4	1617	20	AA135617

12	742.5	8.3	1132	20	AA135048	Chlamydia pneumoniae
13	685.5	7.7	1530	21	AA13635	C. trachomatis pmp
14	665.5	7.4	930	20	AA135052	Chlamydia pneumoniae
15	660.5	7.4	930	21	AA190240	Chlamydia antigen
16	655.5	7.3	914	20	AA188429	Chlamydia pneumoniae
17	655.5	7.3	928	20	AA188418	Chlamydia pneumoniae
18	652.5	7.3	930	20	AA188424	Chlamydia pneumoniae
19	636	7.1	609	20	AA135161	Chlamydia pneumoniae
20	619.5	6.9	928	21	AA190237	Chlamydia antigen
21	616	6.9	885	21	AA190238	Mature Chlamydia a
22	567	6.3	871	21	AA195550	Chlamydia pneumoniae
23	560	6.3	928	20	AA188421	Chlamydia pneumoniae
24	559	6.3	928	21	AA194327	Chlamydia pneumoniae
25	555.5	6.2	925	21	AA199843	Chlamydia pneumoniae
26	555.5	6.2	936	21	AA199842	Chlamydia pneumoniae
27	555	6.2	375	21	AA195549	Chlamydia pneumoniae
28	551.5	6.2	931	21	AA192715	Chlamydia pneumoniae
29	529.5	5.9	922	20	AA188419	Chlamydia pneumoniae
30	525.5	5.9	922	21	AA195548	Chlamydia pneumoniae
31	524.5	5.9	918	21	AA169369	Amino acid sequenc
32	523.5	5.9	918	20	AA188422	Chlamydia pneumoniae
33	522.5	5.8	922	20	AA134597	Chlamydia pneumoniae
34	517.5	5.8	949	20	AA135060	Chlamydia pneumoniae
35	516.5	5.8	928	20	AA188417	Chlamydia pneumoniae
36	516.5	5.8	928	21	AA190236	Chlamydia antigen
37	506	5.7	841	21	AA192818	C. pneumoniae CPN1
38	504	5.6	841	20	AA188420	Chlamydia pneumoniae
39	498	5.6	1013	20	AA16738	C. trachomatis F s
40	496.5	5.6	643	20	AA135056	Chlamydia pneumoniae
41	492	5.5	928	20	AA188423	Chlamydia pneumoniae
42	491	5.5	1013	20	AA16737	C. trachomatis B s
43	490	5.5	928	21	AA190239	Chlamydia antigen
44	489	5.5	962	21	AA13633	C. trachomatis pmp
45	485	5.4	1006	21	AA13639	C. trachomatis pmp

#### ALIGNMENTS

RESULT 1	
ID	AA13637 standard; Protein: 1752 AA.
AC	AA13637:
XX	
DT	02-FEB-2001 (first entry)
XX	
DE	C. trachomatis pmp gene protein.
XX	
KW	Chlamydial infection; sexually transmitted disease;
KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW	trachoma; blindness; acute respiratory tract infection;
KW	atherosclerosis; coronary heart disease; antibacterial.
XX	
OS	Chlamydia trachomatis.
XX	
PN	W0200034483-A2.
XX	
PD	15-JUN-2000.
XX	
PF	08-DEC-1999; 99WO-US29012.
XX	
PR	08-DEC-1998; 98US-0208277.
PR	08-APR-1999; 99US-0288594.
PR	01-OCT-1999; 99US-0410568.
PR	22-OCT-1999; 99US-0426571.
XX	
PA	(CORT-) CORTXA CORP.
XX	
PI	Probst P, Bhatia A, Skelky YAW, Fling SP, Jen S, Stromberg EJ;
DR	WPI; 2000-431303/37.
XX	





DT 02-FEB-2001 (first entry)  
DE C. trachomatis pmpC gene protein.  
XX  
XX Chlamydial infection; sexually transmitted disease;  
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
KW trachoma; blindness; acute respiratory tract infection;  
KW atherosclerosis; coronary heart disease; antibacterial.  
XX  
XX Chlamydia trachomatis.  
XX  
XX WO200034483-A2.  
PN 15-JUN-2000.  
PD  
XX  
XX 08-DEC-1999; 99WO-US29012.  
XX  
XX 08-DEC-1999; 98US-0208277.  
PR 08-APR-1999; 99US-0288594.  
PR 01-OCT-1999; 99US-0410568.  
PR 22-OCT-1999; 99US-0426571.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
PI  
XX WPI; 2000-431303/37.  
DR  
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
PT comprises immunogenic portion of Chlamydia antigen, which comprises  
PT amino acid sequence encoded by polynucleotide sequence -  
XX  
XX Claim 2; Pages 190-194; 256pp; English.  
PS  
XX The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamydiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a protein isolated in the  
CC present invention.  
XX  
SQ Sequence 1776 AA;

Query Match 40.2%; Score 3592; DB 21; Length 1776;  
Best Local Similarity 40.5%; Pred. No. 4.le-174;  
Matches 820; Conservative 264; Mismatches 415; Indels 528; Gaps 40;  
QY 1 MKWLSATVFAVPLPSVSGFCPEPK- ---ELNFSRVETSSSTTTTETIGEAGAEIVSG 56  
DB 3 mkfmsatavfaivssvteassiqdqikntdcnsvkvgystsqafdmmladnteyraad 62  
QY 57 NASFTKFTNIPDTTPTNSNSSSSGSETASVSDSDSTTTTPDPKGGAFYNHAHSGVL 116  
DB 63 svsfydf-----stsglprkhlssseasptt-----egv- 93  
QY 117 SPWTRSGTGLTLSEIKMTGEGGAIFSGOCELLFTDLTSLTQNLSQLSGGAIFGGSTI 176  
DB 94 ----- 93  
QY 177 SLSGITKATFSCNAEVPAPVKPTPEKAQTASSETSGSSSGNDSSVSPSSRAEPAA 236  
DB 94 -----ssssggntcensqds----- 108  
QY 237 NLQSHFICATATPAAQTDTTETPTPSHKPGSGGAIYAKGLDITADSO----- 282

DB 109 -----apssgetdkkt---eeeldnggiiyareklitisesqdslnpsielhdns 155  
QY 283 -----EVLFSINKATKGGAIFAEKDYSEFNITSLKVTQNGAEEKGGAIYAKGDLSTQS 336  
DB 156 fffgegevfldhrvalknngaiygekvfenikellvevnisvekgvsgvyakervsien 215  
QY 337 SKQSIFNSNYSKQGGALYVEGGINFQDLEIRIKYNKAGTETPKKI----- 383  
DB 216 vteaufssnggeggggyiseqdmliiscnnvhfgnaagataavkqciideemivlltecv 275  
QY 384 -----TLPSLKAQASAGNADAWASSPQSGSGATTVSDDSGSSGSDSTSETVPV 434  
DB 276 dselsedtlidstpeteqtknngpqd-----ssetkdqvsespestpspd-----v 322  
QY 435 TAKGGGLYTDKNLSITNITGIIETIANNKATDVGGAYVKGTLTCENSHRLQLKLNSSDKQ 494  
DB 323 lgkgggiytekaltitgtidfvsnlatdsgavftkeniscntnslqfknksag-- 380  
QY 495 GGGIYGEDNITLSNLTGKTLFOENTAKEBGGGLFIKGTDKALMTGTLDSFCLINNTSEKH 554  
DB 381 -----qh 382  
QY 555 GGAFTVTRKISQT-YTSDVETIPGTPVHGTVITCNKSTGCGGGVCTKRLALSNLOSI 613  
DB 383 gggayvtqmsvntntsesittp---plvgevifsen-takghgggictnkislslkvtv 438  
QY 614 SISGNSAENGGAHTCPDFFADTAEQ--PAASAAATSTPK---SAPVSTALSTPSSS 668  
DB 439 titksakesgaiftdlasiptdtpesstpsssa-stpevvasakinfraftaeap 497  
QY 669 TVSSLTLLAASSQASPATSNKQTPADTLLIDYVVDTTISKNT-AKKGGGIYAKKAK 727  
DB 498 aapslt-----easdqtdqtdetsndsidviseniinvaingntsakkgaiyggkak 552  
QY 728 MSRIDQLNISEATEIGGICCKESLELDALVLSVTENLYGKEGGGLHAKHTVSNISLK 787  
DB 553 lsrinnlelsgnsgdqvggclitesvefdalgsllshynsaakeggvthskvtlslnk 612  
QY 788 SGFSF-----SNNKANSSST----- 802  
DB 613 stftfadntvkaivestpeapeeippvegeestatenpnsntegssantnlegsgdtd 672  
QY 803 ----- 802  
DB 673 tdtgvvnesqtdsdtnaesgeqlqdstqsnentlpsnsidqsnentdesdshtsei 732  
QY 803 -----GVATTASAP-----AAAAAALQAAAAAAPSSPA----- 830  
DB 733 tdesvsssksgstpqdggaaagsgpqsanacilaksyaaastdspsvsnsgsdv 792  
QY 831 -----TPYSGVVGAIYGEKVTFFSQCSGTCOF 858  
DB 793 tassndpsssgsdgagdsgeptegagsttetptlig--ggaiygetvkiensfgqgif 850  
QY 859 SGNQAIIDNNP--SQSLNVQGGAIYAKTSLTSGSSDAGTSYIFSGNSVTGKSQTQGIA 916  
DB 851 sgnkaidntegssksknlvggavayaktlfnldsgsrtrtvtfsgntvss--qsttgqva 908  
QY 917 GGAIYSPVTNLNCPATFSNNTASIATPKTSSBDGSGNSIKDTIGGAIAGT-AITLSGVYS 975  
DB 909 ggaiysptvtiatpvpvfknsa-----cnnannatdtkrdtfggaigatsavsllsga 962  
QY 976 RPSGNTADLGAAGIATLANANTPSATSGNSITEKTLLENGSFIFERNKRGAIYSFS 1035  
DB 963 hflenvadlgsaiglvpd-----tqn--tetvklesgsvyfeknkalkratyapv 1011  
QY 1036 VSIKGNITFNQNTSHDGSALYFTKDAIESLGSLVLTGNNVTATQASSATSGQNTFTA 1095  
DB 1012 vsikaytafnqnrslleegsaiyftkeasieslgsvlftgnlvtp-lstttegtpatis 1070  
QY 1096 ----NYGAIFGDPGTTQSSQTDAL-LTLLASSGNTFFSNNSLQNNQGDTPASKFCSIAG 1150  
DB 1096 ----- 1150

Db 1071 gdvtkygaalfigdngsqtdnlpklkiasggnicfrnneyrstsdtgtstfcsiaq 1130  
QY 1151 YVKLSQAAGKTIISFDVCVHTSTKTKTGSTQNVYETLDINKENS---NPYTGTVFSS 1206  
Db 1131 dvkltnqaagkgtisfddairtstktgtqataydldinksedsetvnsafgtgtilfss 1190  
QY 1207 ELHENSYPQNAIHLNGTLVLEKTELHVVSPEQKESKLIMEPQAVLSNQNIANGALA 1266  
Db 1191 elhenksyipqnvvlhsgslvlpntelhvlsfeqegsalvntpgvlsnqtvadgalv 1250  
QY 1267 INGLTTDLSSM---GTPQAGEIFSPPELRIVATTSASGG----- 1303  
Db 1251 inmtldlssveknglae-gniftppelriidttsgsggtptdsesnqnsdtkceqnn 1309  
QY 1304 -----SCVSSSIP-----TNPKRISAAPVSGSAATTP---TMSENKVFUTGDLTLI 1346  
Db 1310 ndasngesangssspavaaahsrtrnfaaataatpttptatttsnqvlggikil 1369  
QY 1347 DPNGNFYQNPMLGSLDVLPLIKLPTNTSDVQVYDLTSLGDLFPQKGYMGFTWLDSPN-QT 1405  
Db 1370 dpgnttfgnparlsdqisllvlpdsskmqgkivltgdiapqkgtgttltldpqlqn 1429  
QY 1406 GKLOARWTFEDTVRWVYIPRDNHFYANSILGSONSMIVVKQGLINMLNNAAREDDIAYNN 1465  
Db 1430 gtisalwktdsyrfqwyvprdnhfyansilgsmnvtvkqglndkmlarfdesynn 1489  
QY 1466 FWVSGVGTFLAQOGTPLSEEFYSYSRGTSVAIDAKPRQDFILGAAPSKIVGKTKAIKKMH 1525  
Db 1490 lwisgltnlsqvgtpseefyysrgasvaldakpahdivgaafsmigktsikren 1549  
QY 1526 NYPHKSEYSYQASVYGGKFLPFLNKNQHWALPFLIQGVVSYGHIKHDTTTLTYPSTIHER 1585  
Db 1550 nythkgeysyqasvygkpfhfvlnktekslpillqgvisygyikhdtvthyptir 1609  
QY 1586 NKGWEDLGLADLRISMDLKEPSKSSKRITVYGELEYSIRKQFTEDIDYPRFDDC 1645  
Db 1610 nggewedlglwltairvssvlrtpagqdkritvygeleysirkrqfteteypdryfnc 1669  
QY 1646 AYRNLSLPVGCVEGAIMNCNIMNLKALAYMPSYIRNPNVCKYRVLSSNEAGQVIGCV 1705  
Db 1670 tyrnlaiplmglafegelesgndlimyrfsvaympsiyrnsptckygvlsgegeilcgv 1729  
QY 1706 PRTSARAYSTQLYLGPFWTLYGNVTIDVGMVTLTSLQMTSCGARMITF 1752  
Db 1730 ptrnsargestqlvypplwtlygsytieadahtlahmmncgarmtf 1776

RESULT 4  
AAB13643  
ID AAB13643 standard; Protein; 948 AA.  
XX  
AC AAB13643;  
XX  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE C. trachomatis pmpC gene carboxy terminus protein.  
XX  
KW Chlamydial infection; sexually transmitted disease;  
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
KW trachoma; blindness; acute respiratory tract infection;  
KW atherosclerosis; coronary heart disease; antibacterial.  
XX  
OS Chlamydia trachomatis.  
XX  
XX  
PN WO200034483-A2.  
XX  
XX 15-JUN-2000.  
PD  
XX  
XX 08-DEC-1999; 99WO-US29012.  
PF  
XX  
PR 08-DEC-1998; 98US-0208277.  
PR 08-APR-1999; 99US-0288594.  
PR 01-OCT-1999; 99US-0410568.

PR 22-OCT-1999; 99US-0426571.  
XX  
PA (CORI-) CORIXA CORP.  
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
XX  
XX WPI; 2000-431303/37.  
XX  
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
PT comprises immunogenic portion of Chlamydia antigen, which comprises  
PT amino acid sequence encoded by polynucleotide sequence -  
XX  
PS Claim 2; Pages 216-218; 256pp; English.  
XX  
XX The present invention relates to new nucleic acid sequences and the  
CC proteins encoded by the nucleic acid sequences. The encoded proteins  
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
CC infection. Chlamydiae are intracellular bacterial pathogens that are  
CC responsible for a wide variety of human infections. C. trachomatis  
CC infection is one of the most common sexually transmitted diseases and can  
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
CC the leading cause of preventable blindness worldwide. C. pneumonia is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a protein isolated in the  
CC present invention.  
XX  
XX Sequence 948 AA;  
SQ  
Query Match 30.2%; Score 2703; DB 21; Length 948;  
Best Local Similarity 56.6%; Pred. No. 2.5e-129;  
Matches 543; Conservative 138; Mismatches 204; Indels 74; Gaps 16;  
QY 847 VTFSCQSGTCQFSGNOAIDNNP--SQSSLNVGGAGIAYAKTSLSGSDAGTSYIFSCNSV 904  
Db 11 vkienfsgggifsgnkaidntteggssksnvlggavyakltfndlsgsrrtvtfgntv 70  
QY 905 STCKSQTGGIAGIAGYSPVTILNCIPATFSNNTASTATPKTSSSDSGSNIKDTIGAI 964  
Db 71 ss--qsttgqvaggaityptvtlatpvyfksna-----tnnannatdtkrkdftfgal 122  
QY 965 AGT-AITLGSVRSFGNTADLGAAGTFLANANTPSATSGSONSTIKETLNGSFIIFERN 1023  
Db 123 gatsavsls99gahflenvadlgsaiglvpd-----tqn--tctvklesgyyfekn 171  
QY 1024 QANKGAIYSPSVSIRKNNITFNQNTSTHDGSAIYTKDATIESLGSLVFTGNNVTATQA 1083  
Db 172 kalkratiyapvvsikaytatfnqnrslseegsaifyfckeasieslgsvlftgnlvtpt-1 230  
QY 1084 SSATSGQNTNTA---NYGNAIIFGDPGTQSSQTDAL-LTLASSGNITFSNNSLQNNQ 1138  
Db 231 sttetedpattsgdvtkygaalfigdngsqtdnlpklkiasggnicfrnneyrftss 290  
QY 1139 DTPASKFCISIAGVYKLSLOAAKGTISFFDCVHTSTKTKTGSTQNVYETLDINKENS--- 1195  
Db 291 dtgtsfcsiadgdkltmqaaakgtisfddairtstktgtqataydldinksedsetv 350  
QY 1196 -NPYTGTVFSSSELHENKSYIPQNAIHLNGTLVLEKTELHVVSPEQKESKLIMEPQAV 1254  
Db 351 nsaftgtlfselhenksyipqnvvlhsgslvlpntelhvlsfeqegsalvntpgv 410  
QY 1255 LSNQNIANGALAINGLTIDLSSM---GTPQAGEIFSPPELRIVATTSASGG----- 1303  
Db 411 lsnqtvadgalvinmtdlssveknglae-gniftppelriidttsgsggtptdses 469  
QY 1304 -----GVSSSIP-----TNPKRISAAPVSGSAATTP---TMSE 1334  
Db 470 nqnsddtkceqnnndasngesangssspavaaahsrtrnfaaataatpttptattts 529  
QY 1335 NKVFLTGLTLTLDIPNGNFYQNPMLGSLDVLPLIKLPTNTSDVQVYDLTSLGDLFPQKGYM 1394



```
Db      881 ahmmcgarmif 892
RESULT  6
ID      AAY37828
XX      AAY37828 standard; Protein; 568 AA.
AC      AAY37828;
XX      07-OCT-1999 (first entry)
XX      Chlamydia trachomatis surface exposed protein.
XX      Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW      paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
KW      nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW      bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX      Chlamydia trachomatis.
OS      Chlamydia trachomatis.
XX      WO9928475-A2.
XX      10-JUN-1999.
XX      27-NOV-1998; 98WO-IB01939.
XX      04-NOV-1998; 98US-0107077.
XX      28-NOV-1997; 97FR-0015041.
XX      17-DEC-1997; 97FR-0016034.
XX      (GEST ) GENSET.
XX      Griffais R;
XX      WPI; 1999-371125/31.
XX      Genome sequence of Chlamydia trachomatis
XX      Disclosure; Page 1399; 1755pp; English.
XX      AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC      of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC      vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC      can also be used to control growth of the microorganism. Chlamydia
CC      trachomatis is responsible for a large number of diseases, e.g. eye
CC      diseases such as conventional trachoma, nonendemic trachoma,
CC      paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC      nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC      perihhepatitis, Bartholinitis; pneumopathy in breast feeding infants;
CC      and venereal lymphogranulomatosis. The polypeptides of the invention
CC      may be of use in treating these diseases.
XX      Sequence 568 AA;
SQ      Query Match      28.7%; Score 2568.5; DB 20; Length 568;
      Best Local Similarity 93.4%; Pred. No. 8e-123;
      Matches 512; Conservative 14; Mismatches 21; Indels 1; Gaps 1;
QY      15 PSVSGCFPEPKELNFSRVETSSSTFTTIGEAGAEYIVSGNASFTKTNPTTDTTTP 74
      |  ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      14 pqfqqfasqnlxxlnfrvgtssstftctevgeagaeyivssnasfktfnptnttpp 73
QY      75 TNSNSSSSGGETASVSESDSTTTPDPKGGAFYNAHSGVLSFWTRSCGSLTISEIK 134
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      74 tnsnssnrcasvksdsdtttcdpdkggafynahsgvlsfmcrcstegsiltiseik 133
QY      135 MTGEGAFPSQGLLFTDLTSLTIQNNLSQSGGAIFGGSITSLSGITKATFSCNSAEVP 194
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      134 mtgeggafisqgellftdltgtiqnnlsqsggaifggstislsgitkatsnsaevp 193
QY      195 AVPKPTPEKAQTASSETSGSSSSGNDVSVPSSSSRAEPAANLQSHFTCATPAAQTD 254
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
Db      194 apvkptkpeaqtagetsgsssssgndsssssskaepaaanlqshfcatatpaugd 253
QY      255 TETSTPSHRPGSGGAIYAKGDLTIADSQEVLFSINKATKGGGAIFAEDKDVSFENITSKV 314
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      254 tetstpsrhpgsggaiyakgdltiadskvlfinsinkatdkgggaifaekdvsvfenitalkv 313
QY      315 QTNGAEKGGAIYAKGDLSTQSSKQSLFNSNYSKQGGGALYVEGGINFQDLEIRIKYNK 374
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      314 qtngaeekggaiyakgdlslqssksqslfnsnyskqgggalyvegninfqdleerikynk 373
QY      375 AGTFETTKITLPSLKAQASAGNADAWASSPQSGGATTVSDSGDSSGSDSTSTVPV 434
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      374 agtfkcktktlps-kagasagnadawasspqsssgattvsgdsgssgnsdtktpv 432
QY      435 TAKGGGLYTDKNLSINFTIETIANNKATDVGGGAYVKGTLTCENSHRLOPLKNSDDKQ 494
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      433 takggglytdknlslcniitgiteiainkatdv99gyvkgtiltcknshrlqlfknssdkq 492
QY      495 GGGIYGEDNITLSNLTGKTLFQENTAKEEGGLFIKGTDKALTMGTGLDSFCLINNTSEKH 554
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      493 ggggygednltlsnltgktilfgentakkegggllfkgtdkalmtgldsfcllnntsekh 552
QY      555 GGGAFVTK 562
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      553 gggafvtq 560
RESULT  7
AAY35162
ID      AAY35162 standard; Protein; 1146 AA.
XX      AAY35162;
XX      13-SEP-1999 (first entry)
XX      Chlamydia pneumoniae cellular envelope protein.
XX      Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW      sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope.
XX      Chlamydia pneumoniae.
XX      WO9927105-A2.
XX      03-JUN-1999.
XX      20-NOV-1998; 98WO-IB01890.
XX      04-NOV-1998; 98US-0107078.
XX      21-NOV-1997; 97FR-0014673.
XX      (GEST ) GENSET.
XX      Griffais R;
XX      WPI; 1999-357842/30.
XX      Genome sequence of Chlamydia pneumoniae
XX      Page 1027-1029; Disclosure; 1912pp; English.
XX      AAY34584-Y35879 represent the proteins encoded by all the open reading
CC      frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC      C. pneumoniae causes respiratory disease such as pneumonia and
CC      bronchitis and is thought to be a contributing factor in heart
CC      disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC      nodosum or pharyngitis. The polypeptides encoded by the open reading
CC      frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC      immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC      nucleotides sequences can also be used as immunogenic compositions,
CC      especially where the vector directs the expression of a neutralising
CC      epitope of C. pneumoniae.
```



CC It is an example of C. pneumoniae antigenic polypeptides of the  
CC invention (see AA9543-55) that are encoded by open reading frames  
CC (see AAA50030-42) identified in the C. pneumoniae genome. The  
CC polypeptides are useful in the diagnosis, treatment and prevention  
CC of Chlamydia infection. They can be prepared by recombinant  
CC methods using transformed unicellular host cells. The  
CC polypeptides, or immunogenic fragments of them, or fusion proteins  
CC that include an additional polypeptide such as a heterologous  
CC signal peptide or a polypeptide having adjuvant activity, are  
CC utilised as vaccines to protect against Chlamydia infection.  
CC The polypeptides are also used in diagnostic kits and in methods  
CC of detecting Chlamydia infection.  
XX  
SQ Sequence 469 AA;

Query Match 14.2%; Score 1272.5; DB 21; Length 469;  
Best Local Similarity 47.3%; Pred. No. 4.3e-57;  
Matches 239; Conservative 89; Mismatches 140; Indels 37; Gaps 3;

QY 1249 MPPGAVLSNQNTANGALAINGLTIDLSMGTQPAQGEIFSPPELRIVATVTSASGGSGVSS 1308  
Db 1 mppgsvlnshksaegaiainnvildseivptkdnatvapptklvstrnad----- 52

QY 1309 SIPTNPKRISAAVPSGSAATPTMSKNVFLGDLTLPDNGNFYONPMLGSDLDVPLIK 1368  
Db 53 -----skkdiitgtvtlldpngnlyqnsylgedrditlfn 88

QY 1369 LPTNTSD-VQVYDLTSLGDLFPQKQVMGTWTLDSNPQTKLOARWTFDNRWVYIPRON 1427  
Db 89 idnsasgavtactnvtlqgnlgakkgyigtwnldpnssgskllkwtfdkylrwpypidn 148

QY 1428 HFYANSILGSQNSMIWVKGGLNNMLNNAFDDIAYNNFWSGVGTFLAQOQTPLSEFS 1487  
Db 149 hfynsiwgaqslvtvngilgmnnarfedpafnnfwsaigsflrkevsrnsdft 208

QY 1488 YVSRGTSVAIDAKPRQDFILGAFAFSKIVGKTKRAIKKMHNFYHKGSEYSYQOASVYGGKFLY 1547  
Db 209 yhgrytaavdakarqrefilgaafsqvfghaeseyhldnykkgshstqaslyagnify 268

QY 1548 FLLNKHQHWALPFLQGVVSYGHKIHDTTLYPSIHERNKGWEDLGWLADLRISMDLKE 1607  
Db 269 fpairsr-----piligvfatygmqndtttypsaeknmanwdsiawldrfsvdlke 324

QY 1608 PSKDSKRITVCELEYSSIROKQTEIDYDPRHFDDCAYRNLSLPVGCAGEAIMNCNI 1667  
Db 325 paphstarltfyeaeaytirrkeftelqyqrsfacsacyglaipgtgfsvdgalawrei 384

QY 1668 LMYNKALAYMPSIYRNPNVCKYRVLSSNEAGOVICGVPTRTSARAEYSTQLYGPFWTL 1727  
Db 385 ilynkvsaaylpvlirnpkayevlstkkgvnnvnlptrnaaraevssqilylgsywtl 444

QY 1728 YGNYTIDVGMVTLTSMQTCGARMIF 1752  
Db 445 ygcytidasmntlvqmganggirfvf 469

RESULT 9

AAV37823

ID AAV37823 standard; Protein; 915 AA.

XX

AC AAV37823;

XX

DT 07-OCT-1999 (first entry)

XX

DE Chlamydia trachomatis cellular envelope protein.

XX

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX

OS Chlamydia trachomatis.

XX WO9928475-A2.  
PN 10-JUN-1999.  
PD 27-NOV-1998; 98WO-IB01939.  
XX 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.  
XX (GEST ) GENSET.  
XX Griffais R;  
PI WPI; 1999-371125/31.  
XX Genome sequence of Chlamydia trachomatis  
PT Disclosure; Page 1391-1393; 1755pp; English.  
XX AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nonendemic trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC perihapatitis, bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.  
XX Sequence 915 AA;  
SQ

Query Match 11.2%; Score 1002; DB 20; Length 915;  
Best Local Similarity 27.3%; Pred. No. 6.3e-43;  
Matches 331; Conservative 153; Mismatches 323; Indels 404; Gaps 34;

QY 1 MWLSATAPAAVLPSVSGFCPEPK-----ELNFSRVETSSSTTTTETIGEACAEIVSG 56  
Db 2 mkflsatavfaavllsvtassiqdqikntdcnvskvgystsqafdmldadnteayraad 61

QY 57 NASFTKFTNPTDPTTPTNSNSSSGTASVSDSTTTTPDPKGGAFNAHSGVL 116  
Db 62 svsfydf-----stsglprkhlssseasptt-----egv- 92

QY 117 SPMTRSGTEGSLTSEIKMTGEGGAIFSQGELLFTDLTSLTIQNLSQLSGGAIFGGSTI 176  
Db 93 -----ssssgentensqds----- 92

QY 177 SLSGITKATFSCNSAEVPAPVKKPTPEKQATASSETSGSSSSSGNDSDVSSPSSRAEPAA 236  
Db 93 -----ssssgentensqds----- 107

QY 237 NLQSHFICATATPAAOTDTETSPSHKPGGGAIVAKGDLTITADSO----- 282  
Db 108 -----apssgetdkkt-----eeeldnggiyaiyekltisesqdslnpsielhdns 154

QY 283 -----EVLFSINKATKGGGAIFAEXKDVSEFNITSLKVQTNGAEEKGGAIVAKGDLISQS 336  
Db 155 fffgegevfidhrvalknggalygekvfvenlksllvevnisxekgsvayakervslen 214

QY 337 SKQSLFNSNYKQGGGALYVEGINFQDLEETRIKYNKAGTFTETKXI----- 383  
Db 215 vteatfssnggeggggilyseqdmliisdcnvfhfgnaagatavkqcldeemilvitacv 274

QY 384 -----TLPSLKAQAAGNADAWASSPQSGGATTVSDGSSSDSDSTSTVTV 434  
Db 275 dalsedtldtpetqtknsngqdg-----ssetkdtqvsespesptpdd-----v 321

QY 435 TAKGGGLYTDKNLSITNITGIETIANNKATDVGGGAYVRKGTLTTCENSHRLQLKNSSDKQ 494

Db	322	lgk9gglyteksitigcigtidfvsniaadsgavfckenslcutntslqflnsg--	379
QY	495	GGGIYGEDNITLSNLTKGTLFOENTAKERGGGLFIKGTDKAL/TWTGLDSFCLINNTSEKH	554
Db	380	-----	qh 381
QY	555	GGGAFVTKESIQP-YTSDVETIPGTPYHGEVTITGKNKSTGGNGGGVCTKRLALSLNQSI	613
Db	382	ggdayvtqmtsvnttssesittp--plvgevifsen-takghggigctnklslslnktv	437
QY	614	SIGNSAAENGGAHCTCPDSFPDTADTAEQPAAASAATSTPKSAPVSTALSPSSSTVSSL	673
Db	438	tltnsaesggaifdlasipdtptes-----	stpsssspast 477
QY	674	TLIAAS-----SQASPA-----TSNKETODPNADTDLLIDYVVDTTISKNT	714
Db	478	xevvasaklnrfastaepaapsltaeasdqtdtetsdntsdidvsielnlvaingnt	537
QY	715	-AKGGGIVYAKKAKMRIDQLNISENSATEIGGGICCKESLELDALYSLVSTENLVGKEG	773
Db	538	sakkgdalygkkaaklrinnlelsgnssqdvgggicltsevefdaigslshynsaakeg	597
QY	774	GGHAKTVNTLSNLKSGFSFSNN--KANSSSTGVATTASAPAAAAASLQAAAAAAPS----	827
Db	598	gvihskvtlsnlksftfadntvkaivestpeapeeiipreve--estatenpnnte	654
QY	828	-SPATPTYSVWG-----GAIYEKVTFSOCSTCFQSCNQAIIDNNPSSQLNVGGGA	879
Db	655	gsaantnlegsgqdtadtgtgvvnnesqdtst-tgnae-sgeqlqdstqneentlps	712
QY	880	IYAKTSLSIGSDAGTSYIFSGNSVSTGKSQTTGQIAGAIYSPVTVTLNCPA---TFSNN	936
Db	713	idqsnentdessahteeitdesvssksgstpqdgaass-----gapsgqgsisan	767
QY	937	TASTATPKTSSGDSGNSIKDTIGGAIAGTATILSGVSRFSGNTADUGAAIGTLANANT	996
Db	768	aclaksyaastdspsvns-----sgsdvtassdnpdsssgd-----sagds	810
QY	997	PSATSGGSONITEKITLENGSPFIERNQAKRGAIYSPSVTIKGNITFNONTSTHDGSA	1056
Db	811	egptepeagstetctpllg-----galygetvki-----enfs-----	844
QY	1057	IYFTKDATIESLGSVLFTGNNVTATQASSATSGQNTNTANYGAAIFG-----DPGT-TQ	1109
Db	845	-----gggifsgnka-----idnttegssksnvlvggavyaktlfnldsgtlde	888
QY	1110	SSQTDAILTLUL 1120	
Db	889	lspspgillsll 899	
RESULT	10		
AAB13644			
ID	AAB13644	standard; Protein; 821 AA.	
XX			
AC	AAB13644;		
DT			
XX	02-FEB-2001	(first entry)	
DE			
XX			
DE			
KW	C. trachomatis pmpC gene amino terminus minus signal sequence protein.		
KW	Chlamydial infection; sexually transmitted disease;		
KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;		
KW	trachoma; blindness; acute respiratory tract infection;		
XX	atherosclerosis; coronary heart disease; antibacterial.		
OS	Chlamydia trachomatis.		
XX			
PN	WO200034483-A2.		
XX			
PD	15-JUN-2000.		





```
Db 306 dkgtglaisqnmqsfnsnttttanggaiaitkctldnttltfdqnta-----tagc 356
Qy 659 STALSTPSSSTVSSLTLLAASSQASPATSKNETQDPNADTDLIDYVDTTTSKNTAKKG 718
Db 357 ggaityt-----etedfs-----lkgstglvtfstntakg 386
Qy 719 GGIYAKKAKMSRIDQINISENATEIGGICCKESLELDALVLSVTENLVNKEGGGLHA 778
Db 387 galyk-----ens-----slgtntnll-----404
Qy 779 KTVNISNLKGFSGFNKANSSTGVATTASAPAAAAAASLOAAAAAPSSPATPTYSVV 838
Db 405 -----fsgnktgpnssanegcggallsflesasvt-----438
Qy 839 GGAIYGEKVTFQSCGTCGTFSGNQAIDNNPSSQSLNVQGGAIYAKTSLSIGSSDAGTSYI 898
Db 439 -----kglwiednevnslgnta-----cvsggaiyatkalhgn-----ttilt 478
Qy 899 FSGNSVSTGKSQTTQGIAGAIYSPTVTLNC-----PATFSNNTASIAPTKSSSEDCSSG 953
Db 479 fdgn-----taetagaiaiyettedflltgstgtvtfstnta-----ktagalhthg 524
Qy 954 NSIKDPTIGGAIAGTATLLTSGVSFRSGNTADLGAAIGTLANANTPSATSGQNSI-----1007
Db 525 n-----tsftkhalvfsgnsa-----tatatttdqecgcgallcnise 564
Qy 1008 ----TEKITL-ENGSPIFFERNOANKR-GAIYSPSVSIKGN-NITFNQNTSTHDGSAIYFT 1060
Db 565 sdiatkslittenesisfinttakrsggiyapkcvigsesinfngntaetsggaiaiy-s 623
Qy 1061 KQATIESLGSVLTGNNVTATQASSATSGQNTNTANYGAIFGDPGTTQSSQTDAILTL 1120
Db 624 kalsitangpvsftnn-----sgkg-----gaiyiadsge-----lsle 658
Qy 1121 ASSGNITFSNNSLQNNQDTPASKFCISIAGYVKLSLOAAKGKTIISFDCVHSTKTKGST 1180
Db 659 aldgdlfsgn-----rategstpnshlgagakitkilaapghgtiifydptmeapasgt 716
Qy 1181 QNVYETLDIN-----KEEN-----SNPYTGTVIFSS-ELHENKSYIP 1216
Db 717 ---ieelvinpvrkaivppqpkngpiasvrvvpvapnptgtlvfssgkklpsqdasip 773
Qy 1217 QNAI-----LHNGTILVLEKTELHVYVFPQKESKLIMEPGAVL--SNQNIANGALA 1266
Db 774 antttlndqkinlaggnvvlkegatiqvysftqgpdstvmdagttletttntntdgsid 833
Qy 1267 INGLTIDLSSMGTPOAGEIFSPPELRIVATTSASGSGSVSSSIPTNPKRISAAVPSGSA 1326
Db 834 lknslvnldalqgr-----mitiavnstsgglkis-----864
Qy 1327 ATPTMSENKVF LTGDLTLIDPNGFYQNPMLGSDLDVPLIKLPTNTSDVQVYDLR--LS 1384
Db 865 -----gdlkfhnnegsfydnpglkanlnlpfldlsstsgtgnlddfnpi 910
Qy 1385 GDLFPQKGYMGWTWLDNSNPTQC-----KLOARWTFDYR-----RWYIPRONHFFYANSI 1434
Db 911 smaapdygvqgswtl--vpkvagagkvtiavaeqalgytpkpelratlvp-----nsl 961
Qy 1435 LGSQNSMIVVVKGLINLNMLNARFDDIAYNNFWVSGVGFFLAQCGTPLSEEFYSYSRGTS 1494
Db 962 wnayvnihsiqeia-----tamsdapsbgwiigisignafhqdkkenagflisrgyi 1016
Qy 1495 VAIDAKPRODFILGAAFSIVGKTK--AIKKMHNYPHKGS---EYSYQASVYGG---KFL 1546
Db 1017 vggsmtpdqeytfavafslqfkskdyvvsdiksqvayagslcaqssyviplhsrlrhvl 1076
Qy 1547 YFLLNKQHGVALPFLIQGVVSVGHIKHDTTLTTPSPSIHERNKGDWE 1591
Db 1077 skvlpelpg-etplvlhgvsvgrnhhmttklan-ntgkgsdwd 1119
```

RESULT 13  
AAB13635

```
ID AAB13635 standard; Protein; 1530 AA.
XX
AC AAB13635;
XX
DT 02-FEB-2001 (first entry)
XX
DE C. trachomatis pmpD gene protein.
XX
KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial.
XX
OS Chlamydia trachomatis.
XX
PN WO200034483-A2.
XX
PD 15-JUN-2000.
XX
PF 08-DEC-1999; 99WO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
(CORI-) CORIXA CORP.
XX
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
XX
WPI; 2000-431303/37.
XX
Isolated polypeptide for diagnosis and treatment of Chlamydia infection
comprises immunogenic portion of Chlamydia antigen, which comprises
amino acid sequence encoded by polynucleotide sequence
XX
Claim 2; Pages 186-190; 256pp; English.
XX
The present invention relates to new nucleic acid sequences and the
proteins encoded by the nucleic acid sequences. The encoded proteins
comprise an immunogenic portion of a Chlamydia antigen. The encoded
proteins are useful for the serodiagnosis and treatment of Chlamydia
infection. Chlamidia are intracellular bacterial pathogens that are
responsible for a wide variety of human infections. C. trachomatis
infection is one of the most common sexually transmitted diseases and can
lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
and infertility. Trachoma due to ocular infection with C. trachomatis is
the leading cause of preventable blindness worldwide. C. pneumonia is a
major cause of acute respiratory tract infections in humans and is also
thought to play a role in the pathogenesis of atherosclerosis and
coronary heart disease. The present sequence is a protein isolated in the
present invention.
XX
SQ Sequence 1530 AA;
XX
Query Match 7.7%; Score 685.5; DB 21; Length 1530;
Best Local Similarity 21.7%; Pred. No. 1.6e-26;
Matches 417; Conservative 251; Mismatches 670; Indels 583; Gaps 82;
QY 8 AVFAAALPSVSGRC-----FPEPKELNFSRVETSSSTFT 42
Db 17 svvaailasvsglascvdlhagggsvnelvygvpqavllldirdlrvsgksdgaeglyr 76
QY 43 ETIGEAGAEIVVSGNASFTKFTNIPRTDTPPT-----NSNSSSSSGETAS-VSESDSDS 95
Db 77 llvgdpssfqekdadtlpgkveqstlfsvtnpvnvfgvqdgqdvsgglicsfstsnlds 136
QY 96 TTTTPDPKGGAFYNAHSGVLSFPMRTSGSLTLEIKMTGEGGAIFSGOGLLFTDLTS 155
Db 137 -----prdesflg-----iafvgdsskag-iltcdvksalgaalystedlif----- 179
QY 156 LTQNNLSQLSGGAIFGG-STISLSGITKA-----TFSCNSAEVPAPVKPTEPKAQTASE 210
```

Db 180 -----ekikgglefascslleggacaagsilihdcqilgv-----khattavn 223  
Qy 211 TSGSSSS-----GNDVSSPSSRRAEPAANLOSHFICATATPAQAQTDTETSTPSHKP-- 264  
Db 224 aegssandhlgf9ggaffvtlsge-----kslympagdmvvanocgaaisfegnsanf 277  
Qy 265 GSGGAIYAKGD-LTTADSQEVLSTNKATKDGGAIFAEKDVSFENITSLKVQTN---GAE 320  
Db 278 angaiaasgkvlfvandkktfienral-sggaiaasdiacfncacaelvfkgncaigte 336  
Qy 321 EK-----GGAIVAKGDLSTQSKQSLFNYSKQGGALY-----VEGGINFOD----- 364  
Db 337 dkgslggaiaaslgvtlqgnhgictcdknesaggaifgkncqisdnebpvfrdstac 396  
Qy 365 -----LEETRIYKNAG-TFETKKTITLPSLKAQASAGNADAWASSPOSGSGATTVS 415  
Db 397 lgggalaagelvislqnnagaisfegk-----asfgglaacgsfssagga----- 441  
Qy 416 DSGDSSGSDSTSETPVTAKGGGLYTDKNLSITNITGIEIANNKATDVGGGAYVRGT 475  
Db 442 -----svlgtidiskn-----lgaisftrl 462  
Qy 476 LTCENSHRLQFLKNSDKOGGIYGEDNITLSNLTKTLFQENTAKERGGGLFKGTDKA 535  
Db 463 cttslqgmey-----gggalfgenisisenagvtfkdnivk----- 501  
Qy 536 LTMGLDLSFCLINNTSEKHGGGAFVTKESIQSYTSDVETIPGIVPHGETVITGNKSTGG 595  
Db 502 -----tfasngkilgg-----gailatgkveitn 525  
Qy 596 NGGGVCTKRLALNSLQISISGNSAAEGGAHTCPDSFPTADTAEOPAASAAATSTPKS 655  
Db 526 nsgv-----isftgnara-----pqalptqe--efp-----lfskkeg 556  
Qy 656 APVSTALSTPSSSTVSSITLAAASSQASPAFNSKNTQDPNADTDLL-----IDVYVDT 708  
Db 557 rplsgygggailgrevail---hnaavvfeqnlqcseeeatlqccgggavghgmst 613  
Qy 709 TISKNTAKGGGIYAKKAKMSRIDQNLSENSATIGGICCKESLELDALVLSVTENL 768  
Db 614 slvgnsvrfgnnya-----mgvgvggallskvtqlagnsvdfsrni 657  
Qy 769 VGKEGGGLHAKVTNISLNUKSGF-SFSNNKANSSSTGVATTASAPAAAAAASLQAAAAAPS 827  
Db 658 aslggalaqasegncelvdngyvlfrdnrg----- 688  
Qy 828 SPATPTYSGVGGAI---YGEKVTFSCSGTCQPSGNOAI----- 864  
Db 689 -----vyggaiaclrgd-vvisgnkgrvefkdniaatrllyveetvekeveepapeq 738  
Qy 865 -DNNPSQSLNVQGGAIYAKTSLSTIGSDACTSYIFSGNSVSTGKSOTGTQIAGGAIYSP 923  
Db 739 kdnelsflgsveqfllaqalfasegdldl---pessisseeakrrccaggaifak 795  
Qy 924 TVTLNCPATFSNNTASIAPTKTSSEDSGSGNSIKDTIGGAIAGTAGTATILSGVSRFSGNTAD 983  
Db 796 rvri-----vdnqeavfs-----nnfsdiygal-----ftg----- 823  
Qy 984 LGAAI GTLANATPATSQSONSIEKTTLENGSFIERNQANKGAIYSPSVSTKGN-- 1041  
Db 824 -----slreedkld-----gqi--pevlisgnag 845  
Qy 1042 NITFNONTS-----THDGSAIYFKKDATI-ESLGSVLFTGNNTVATQASSATSGQNTNT 1094  
Db 846 dvvifsgnskrdehlpht9gggaicqnlitstqntgnvify-nnv-----acsq-----892  
Qy 1095 ANYGAIFGDPGCTQSSQTDAILTLASSGNITFSNNSLQNNQGDTPASKFCISAGYVKL 1154  
Db 893 ---gavriedhgn-----vllleaaggdivfkgnsfraqg-sdaiytagkeshi-t 938  
Qy 1155 SLOAAKAGTISFFDC-VHTSTKKTGSTONVYETLIDNKEENSNPYTGTVFSSSELHENKS 1213  
Db 1155 SLOAAKAGTISFFDC-VHTSTKKTGSTONVYETLIDNKEENSNPYTGTVFSSSELHENKS 1213

Db 939 alnateghaivfhdalvfenlkerksa-----evlilnsrenpg-ytgsirfl-----eae 989  
Qy 1214 YIPQNAIILHNCTFLVLKTEKTELHVVSFEQEGSKLIMBPGAVLSNQNIANGALAINGLTID 1273  
Db 990 kvpqcihvqqgslellngatlcsgyfkdaqakivlaagsklk----- 1032  
Qy 1274 LSSMGTPOAGEIFPPELRIVATTSSASGGS-----GVSSSIP-TNPKRISAAVPS-GSA 1326  
Db 1033 ildsgtpvqghaiskpeaeiessepegahslwiaknaqtctvpmvdihtisvdlasfss 1092  
Qy 1327 ATPTMSNKVKFLRG-----DLTILDPNGFYQN-PMLGSDLDVPLKLTPTNT--- 1373  
Db 1093 qgegveapgvivp9gsyvrsgelnlelvntgtgyenhalakneakvplmsfvassdea 1152  
Qy 1374 -----SDVQVYDLT-LSGDLFPQGYMGWTWTLDSNPQTGKLOARWTFDYYRRWVY 1422  
Db 1153 saelsnlsvdqlhvateleedyt---ghmgdws-eakiqdgltlvinwnptgyr----- 1204  
Qy 1423 IPRDNHFYANSILGSQNSMIVV-----KQGLTNMNNNAREDDIAYN-----N 1465  
Db 1205 -----lqpqkagalvfnaalweegavlsalknarf---ahnltaqrmeefdystn 1249  
Qy 1466 FWMSGVGTFF-LAQOGTPLSEEFSSYSGTSTVADAKPRQDFILCAAFSKIVGKTKAIKK 1523  
Db 1250 vwgfafgfrtlsaenlvaiddgykgayggasagvdiqimedfivgsgaaflgkmdsqkf 1309  
Qy 1524 MHNYPHKGEYSYQASVYGGKFLYLLNKQHGWALPFLIQGVWSYGHKIKHDTTLLYPSIH 1583  
Db 1310 daevsrkg-----vvgsvyvgfl-----agsvffkqgyslgetcndmkttrygvlg 1354  
Qy 1584 ERNKGDWEDLGLWADLRIS-MDLKEPSKDS--SKRITVYGELEYSSIRQKQFTETIDYDPR 1640  
Db 1355 e-saaswtsrgvadaleveyrslvgvrptfyalhfnpyvevsyasmkfpfteqgrar 1413  
Qy 1641 HFDCCAYNLSLPCCAVEGAIMNCNILMYNKALAYMPSIYRNNPVCYKRYVLSNEAGQ 1700  
Db 1414 sfedasltntitipigmkfelfafikgffsevnslgisayaeayrkevgaqvl---eagf 1470  
Qy 1701 VTCGVPTTSARAESTOLYLGPFWTLVGNVTIDV---GMVT-----LSQWTSQCGARMI 1751  
Db 1471 dweap-mdlprqlrvalentnewssyfstvlgltafcggtstdsklgyeantglrli 1529  
Qy 1752 F 1752  
Db 1530 f 1530  
RESULT 14  
AAY35052  
ID AAY35052 standard; Protein; 930 AA.  
XX  
AC AAY35052;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Chlamydia pneumoniae surface exposed polypeptide.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
OS Chlamydia pneumoniae.  
XX  
PN W09927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
XX 21-NOV-1997; 97FR-0014673.  
XX (GEST ) GENSET.

```
XX Griffiths R;
XX PI
XX DR
XX WPI: 1999-357842/30.
XX PT
XX Genome sequence of Chlamydia pneumoniae
XX PS
XX Page 940-942; Disclosure: 1912pp; English.
XX AA
XX AA34584-Y35879 represent the proteins encoded by all the open reading
XX frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX bronchitis and is thought to be a contributing factor in heart
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX nodosum or pharyngitis. The polypeptides encoded by the open reading
XX frames of the C. pneumoniae genome (see AA34584-Y35879) can be used in
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX nucleotide sequences can also be used as immunogenic compositions,
XX especially where the vector directs the expression of a neutralising
XX epitope of C. pneumoniae.
XX SQ
XX Sequence 930 AA:

Query Match 7.4%; Score 665.5; DB 20; Length 930;
Best Local Similarity 25.0%; Pred. No. 8.1e-26;
Matches 292; Conservative 150; Mismatches 389; Indels 335; Gaps 48;

QY 658 YSTALSTPSSSTVSVSLTLLAASSQASPATS-----NKETQDPNADTDLDDYVVD 707
DB 10 lsstlvtp---llslyatygdaslsptdsfdgaggstftpkstadang-----tnyvlis 61
QY 708 TTRISAKTKGGIVAKRAKMRIDOLNISENSATYIGGGICCKESLELDALVLSLVSTEN 767
DB 62 gnyvindagk-----talgtg-cc-----ftet 83
QY 768 LVKKEGGGLHAKTVINSLNKGFSFNKANSSTGCVATFASAPAAAASLQAAAAAPS 827
DB 84 -----tgdlftfgkgyysfntvdagn-----agaaastadkaltft 122
QY 828 SPATYSGVGVGAIYGEKVTSQC-----SGTCQFSGNQAIIDNNPSQSLNVQGGAI 880
DB 123 gfsnlsfiaapgttvvasgkstlsagalnldngtllsfqns-----neannnggai 175
QY 881 YAKTSLGSSDAGTSYIFSGNSVSTGKSQTTQAGGAIYSPVTVLNCPATFSNNTASI 940
DB 176 takt-lsi-----sgntsitftsnsaklglgais-----saaaisngtqgl 218
QY 941 ATPKTSSEDDSGSGNSIKDITIGGAI---AGTATLTSVSRFSQNTADLGAAGTLANANTP 997
DB 219 vfmnkgetg-----ggalgfaasssitqnsllffsgnta----- 253
QY 998 SATSGQNSITEKITLNGCSFFIFERNQANKRAIY-----SPSVSIKGN-NITFNQNTS 1050
DB 254 -----tdaagkgaiycektgetptltisgnkslftfaens 289
QY 1051 THDGSIAVFTKDATIESLGSVLFTGNVNTATQASSATSQNTNTANYGAAIFGDPCTTQS 1110
DB 290 vtqggai-cahgldisaagptlfsnrc-----gntaagkggaiaiaadsgs--- 334
QY 1111 SOTDAILTLASSGNITFSNNLQNNQGDTPASKFCSAGYVKL-SLQAAKGKTIISFFDC 1169
DB 335 -----lsanqgditflgntlttsaptstrnaiylgssakitnlraaqqgslyfydp 388
QY 1170 VHTSTKTKGSTQNVVETLDINKEENSNP--YTGTVVFSE---LHENK-----SYIPON 1218
DB 389 lasnt--tgas----dvitinqpdsnpldygstivfsgekliadeakaadnftsilkip 442
QY 1219 ATLHNGTLVLKTEKTHLVVSPQKESKLIMEPQAVLSNQNLANGALANGTLIDLSMG 1278
DB 443 lalasgtlalkgnvldvngftqtegtllmqpgtklkad-----teaisitkivvdsl- 498
QY 1279 TPQAGEIFSPPELRIVATTSSASGGSGVSSSIPTNPKRISAAPVSGSAATPTMTSENK-V 1337
```

```
DB 499 -----egnksvs-----ietagaukti 515
QY 1338 FLTGDLTLDTPNGNFYONPMLGSDLDVPLIKLPTNTSDVOVY-DLTLSGDL---FPQKGY 1393
DB 516 tltsplvfqdsngfshntinqaftqplvvftaacaadiyidalltspvtpephlyg 575
QY 1394 MG-----TWTLDSPNQTGKLOARWTFDTY-----RRWVYIPRONHFYAN-----SILG 1436
DB 576 qghweatwadtstaksgtm--twttgynpnperasvvp--dslwasftidrtlqimt 631
QY 1437 SONSMTVVQGLINMLNARFDDIAYNNFVSGVGTFLAQOQTPLUSEEFSYYSRGTSVA 1496
DB 632 sqanslyqqrql-----wasgtanfhhkdksgtnqafrhksygyivg 673
QY 1497 IDAKPRQDFILGAASFKIYKTKAKIKKMHNYPHKSGSEYVQASVY--GGKELYFLLNKHQ 1554
DB 674 gsaeofsenifsvafccqlfgkdkdlfiventsh-----nylaslylqhrafigglpmpsf 728
QY 1555 GWA-----LPFLIOGVSVSYGHKHDHDTTLYPSIHERNKGWEDLGLWADLRISMDLKEP 1608
DB 729 gsitdmkdiplilnaqlsyrkndmdtrytsypea-ggswtnnsgalelgsalalyip 787
QY 1609 SKDSSKRITVYGELEYSSI--RQKQTEIDYDPRHFDDCAYRNLSLPVCAGVEGAINMNCN 1666
DB 788 -keapffggyfplkfqavysrqnfkesgaeaaraafddgdlvncslpvgirl------k 840
QY 1667 ILMYNK---LALAYMPSIYRNPNVCKYRVLSSNEAGQVIGVPTPTSARAEYSTOLYLG 1722
DB 841 lisedeknnfeislalvgdyvrknprsrtslmvsgaswtslcknlraqafilasagshltis 900
QY 1723 PEWTLYGN-----YTIDVGM 1737
DB 901 phvelsgeaayelrgsahinyndcgl 926

RESULT 15
AA34584
ID AA34584 standard; Protein; 930 AA.
XX AA34584;
XX XX
XX 29-AUG-2000 (first entry)
XX Chlamydia antigen CPN100639.
XX DE
XX Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
XX KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
XX KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
XX OS
XX Chlamydia pneumoniae.
XX PN
XX WO200032794-A2.
XX PD
XX 08-JUN-2000.
XX PF
XX 01-DEC-1999; 99WO-CA01147.
XX XX
XX 01-DEC-1998; 98US-0110339.
XX PR
XX 01-DEC-1998; 98US-0110340.
XX PR
XX 01-DEC-1998; 98US-0110427.
XX PR
XX 01-DEC-1998; 98US-0110428.
XX PR
XX 01-DEC-1998; 98US-0110438.
XX XX
XX (CONN-) CONNAUGHT LAB LTD.
XX PA
XX Murdin AD, Oomen RP, Wang J;
XX PI
XX WPI: 2000-412339/35.
XX DR
XX N-PSDB: AAA30853, AAA30854.
XX XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX PT
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: October 25, 2001, 08:39:06 ; Search time 48.31 Seconds  
(without alignments)  
4798.145 Million cell updates/sec

Title: US-09-454-684A-180  
Perfect score: 8942  
Sequence: 1 MKWLSATAVFAAFLPSVSGF.....IDVGMVTLQMTSCGARMIF 1752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	8827.5	98.7	1751	2	084418
2	6488.5	72.6	1672	2	Q9PJY2
3	3550	39.7	1770	2	084419
4	3205	35.8	1460	2	Q9PJY1
5	3112.5	34.8	1723	2	Q92812
6	3111.5	34.8	1723	2	Q9RB59
7	3111.5	34.8	1732	2	Q9K2C1
8	932.5	10.4	1407	2	Q9Z899
9	823.5	9.2	1276	2	Q9JRW2
10	743.5	8.3	1609	2	Q9RB58
11	743.5	8.3	1609	2	Q9Z6U5
12	674	7.5	1531	2	084818
13	665.5	7.4	930	2	Q9RB66
14	661.5	7.4	930	2	Q9Z393
15	655.5	7.3	914	2	086163
16	655.5	7.3	928	2	Q9RB65
17	622.5	7.0	1520	2	Q9PLB0
18	579.5	6.5	926	2	P71135
19	577	6.5	2232	5	P91365

20	567	6.3	847	2	P71132	P71132 chlamydia p
21	563.5	6.3	846	2	P71133	P71133 chlamydia p
22	559	6.3	839	2	P77792	P77792 chlamydia p
23	555.5	6.2	936	2	Q92898	Q92898 chlamydia p
24	553.5	6.2	936	2	Q9JSA2	Q9JSA2 chlamydia p
25	530.5	5.9	922	2	Q929G5	Q929G5 chlamydia p
26	529.5	5.9	922	2	Q924H9	Q924H9 chlamydia p
27	528.5	5.9	922	2	Q9K1Y9	Q9K1Y9 chlamydia p
28	517.5	5.8	949	2	Q9K299	Q9K299 chlamydia p
29	516.5	5.8	928	2	086164	086164 chlamydia p
30	516	5.8	975	2	084417	084417 chlamydia t
31	509	5.7	772	2	Q9RB71	Q9RB71 chlamydia p
32	506	5.7	841	2	Q9Z3A1	Q9Z3A1 chlamydia p
33	504.5	5.6	2586	5	Q9VTK8	Q9VTK8 drosophila
34	504	5.6	1013	2	084879	084879 chlamydia t
35	495	5.5	928	2	Q92398	Q92398 chlamydia p
36	495	5.5	928	2	Q9PJY3	Q9PJY3 chlamydia m
37	470	5.3	986	2	Q92896	Q92896 chlamydia p
38	470	5.3	995	2	Q9K2A1	Q9K2A1 chlamydia p
39	467.5	5.2	3381	2	Q9KX33	Q9KX33 streptococ
40	466	5.2	1016	2	084880	084880 chlamydia t
41	461.5	5.2	978	2	Q9RB63	Q9RB63 chlamydia p
42	459.5	5.1	2178	2	Q9KWR3	Q9KWR3 streptococ
43	456.5	5.1	978	2	Q92895	Q92895 chlamydia p
44	448.5	5.0	987	2	Q9PL45	Q9PL45 chlamydia m
45	446.5	5.0	1275	5	O76602	O76602 caenorhabd

#### ALIGNMENTS

RESULT 1  
084418 PRELIMINARY; PRT: 1751 AA.  
AC 084418;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PUTATIVE OUTER MEMBRANE PROTEIN B.  
GN PMPB.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/CX;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.I., Zhao Q., Koonin E.V., Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis.";  
RL Science 282:754-759(1998);  
DR EMBL; AE001314; AAC68010.1; -;  
DR InterPro; IPR003357; -;  
DR Pfam; PF02385; OMP; 1.  
DR Pfam; PF02415; DUF145; 1.  
SQ SEQUENCE 1751 AA; 183011 MW; C6BE3F44DA7618B3 CRC64;

Query Match 98.7%; Score 8827.5; DB 2; Length 1751;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1732; Conservative 8; Mismatches 10; Indels 3; Gaps 2;

QY 1 MKWLSATAVFAAFLPSVSGFCPEPKELNFSRVETSSSTFTTETIGAGAEYIVSGNASF 60  
Db 1 MKWLSATAVFAAFLPSVSGFCPEPKELNFSRVETSSSTFTTETIGAGAEYIVSGNASF 60  
QY 61 TKFTNPTTDTTPTNSSSSSGSETASVSDSDSTTTTDPKGGGAFYNAHSGVLSFWT 120  
Db 61 TKFTNPTTDTTPTNSSSSSGSETASVSDSDSTTTTDPKGGGAFYNAHSGVLSFWT 120

```
Qy 121 RSGTEGSLTLSEIKMTGREGGAIFSGELLEFTDLTSLATQNNLSQLSGGAIFGGSTISLSG 180
Db 121 RSGTEGSLTLSEIKMTGREGGAIFSGELLEFTDLTGLTQNNLSQLSGGAIFGESTISLSG 180
Qy 181 ITKATFSCNSAEVPAPVKPPEPKAQTASETSGSSSSGNDVSPSSSRAEPAANLQS 240
Db 181 ITKATFSSNSAEVPAPVKPPEPKAQTASETSGSSSSGNDVSPSSSRAEPAANLQS 240
Qy 241 HFICATATPAATQDETETPSHKPGSGGAIYAKGDLTADSOEVLFSINKATKDGGAIFA 300
Db 241 HFICATATPAATQDETETPSHKPGSGGAIYAKGDLTADSOEVLFSINKATKDGGAIFA 300
Qy 301 EKDVSFENITSLKVQTNAGAEKGGAIYAKGDLISQSSKQSLFNSYKQGGGALYVEGGI 360
Db 301 EKDVSFENITSLKVQTNAGAEKGGAIYAKGDLISQSSKQSLFNSYKQGGGALYVEGDI 360
Qy 361 NFQDLEETRIKYNKAGTETTKITLPSLKAQASAGNADAWASSPQSGGATTVSDSGDS 420
Db 361 NFQDLEETRIKYNKAGTETTKITLP--KAQASAGNADAWASSPQSGGATTVSNSGDS 418
Qy 421 SSGSDSDTSETVPVTAKEGGGLYTDKNLSITNITGIIETANKKATDVGGGAYVKGTLTCEN 480
Db 419 SSGSDSDTSEIVPATAKGGGLYTDKNLSITNITGIIETANKKATDVGGGAYVKGTLTCEN 478
Qy 481 SHRLQFLKNSSDKQGGGIYGEDNITLSNLTKTLFQENTAKEEGGLPIKGTDKALMTG 540
Db 479 SHRLQFLKNSSDKQGGGIYGEDNITLSNLTKTLFQENTAKEEGGLPIKGTDKALMTG 538
Qy 541 LDSFCLINNTSEKHGGGAFVKEISQTVSDVETIPGTPVHGETVITGNKSTGGNGGV 600
Db 539 LDSFCLINNTSEKHGGGAFVKEISQTVSDVETIPGTPVHGETVITGNKSTGGNGGV 598
Qy 601 CTKRLALNSLOISISGNSAENGGGATCPDSPPTADTAEPAAAAAATSTPKSAP--VS 659
Db 599 CTKRLALNSLOISISGNSAENGGGATCPDSPPTADTAEPAAAAAATSTPKSAPVS 658
Qy 660 TALSTPSSSTVSSLTLLAASQAQSPATSKETQDPNADTDLIDYVVDTTISKNTAKRG 719
Db 659 TALSTPSSSTVSSLTLLAASQAQSPATSKETQDPNADTDLIDYVVDTTISKNTAKRG 718
Qy 720 GIYAKKAKMSRIDQINISENATEIGGGICCKESLELDALVSLVTENLVKKEGGLHAK 779
Db 719 GIYAKKAKMSRIDQINISENATEIGGGICCKESLELDALVSLVTENLVKKEGGLHAK 778
Qy 780 TVNINSLKGSFNSNNKANSSTGVATTASAPAAAAAASLQAAAAAASPSPATPYSGVVG 839
Db 779 TVNINSLKGSFNSNNKANSSTGVATTASAPAAAAAASLQAAAAAASPSPATPYSGVVG 838
Qy 840 GAIYGEKVTFCQSGTCQFSGNQADNNPSSSLNVQGGAIYAKTSLSIGSDAGTSYIF 899
Db 839 GAIYGEKVTFCQSGTCQFSGNQADNNPSSSLNVQGGAIYAKTSLSIGSDAGTSYIF 898
Qy 900 SGNSVSTGKSTQTTGAIAGGAIYSPVTLNCPATFSNNNTASATPKTSSSEDSSGNSIKDT 959
Db 899 SGNSVSTGKSTQTTGAIAGGAIYSPVTLNCPATFSNNNTASATPKTSSSEDSSGNSIKDT 958
Qy 960 ICGAIGTAGTATLSGVSRESGNTADLGAAGTILANANTPSATSGSONSTIKETILENGSFI 1019
Db 959 ICGAIGTAGTATLSGVSRESGNTADLGAAGTILANANTPSATSGSONSTIKETILENGSFI 1018
Qy 1020 FERNOANKRGAIYSPSVSIKGNITFNQNTSTHDGSAIYFTKDATIESLGSVLFNGNVT 1079
Db 1019 FERNOANKRGAIYSPSVSIKGNITFNQNTSTHDGSAIYFTKDATIESLGSVLFNGNVT 1078
Qy 1080 ATOASSATSGQNTNTANTGAALFGDPGTTQSSQTDAILITLLASSGNTIFSNNSLQNNQGD 1139
Db 1079 ATOASSATSGQNTNTANTGAALFGDPGTTQSSQTDAILITLLASSGNTIFSNNSLQNNQGD 1138
Qy 1140 TPASFCSITAGVVKLSLQAAKGTISFFDCVHTSTPKTGSTQNVYETLDINKEENSYPYT 1199
Db 1139 TPASFCSITAGVVKLSLQAAKGTISFFDCVHTSTPKTGSTQNVYETLDINKEENSYPYT 1198
Qy 1200 GTIVFSELHENKSYIPQNAILHNGTLLVLEKTELHVVSFEQKEGSKLIMBPGAVLSNQ 1259
```

```
Db 1199 GTIVFSELHENKSYIPQNAILHNGTLLVLEKTELHVVSFEQKEGSKLIMBPGAVLSNQ 1258
Qy 1260 IANGALAINGLITDLSMGTPOAGEIFSPPELRIVATSSASGGVSSSTPTPNKRISA 1319
Db 1259 IANGALAINGLITDLSMGTPOAGEIFSPPELRIVATSSASGGVSSSTPTPNKRISA 1318
Qy 1320 AVPSGSAATPTMSKENKVFLLTGLDITLDIPNGFNQNPMLGSDLDVPLIKLPTNTSDVOVY 1379
Db 1319 AAPSGSAATPTMSKENKVFLLTGLDITLDIPNGFNQNPMLGSDLDVPLIKLPTNTSDVOVY 1378
Qy 1380 DLTLSGDLFPQKGYMGWTWLTDSNFPQTKLQARWTFDTYRRWVYIPRDNHFNANSILGSQ 1439
Db 1379 DLTLSGDLFPQKGYMGWTWLTDSNFPQTKLQARWTFDTYRRWVYIPRDNHFNANSILGSQ 1438
Qy 1440 SMIVVVKQGLINMLNNAFFDDIAYNNFWVSGVGFLLAQOQTPLEEFPSYISRGTSVAIDA 1499
Db 1439 SMIVVVKQGLINMLNNAFFDDIAYNNFWVSGVGFLLAQOQTPLEEFPSYISRGTSVAIDA 1498
Qy 1500 KPRODFILGAAFSIVGKTKAIKKMHNYFHKGSEYVQASVYGGKFLYFLNKKHGWALP 1559
Db 1499 KPRODFILGAAFSIVGKTKAIKKMHNYFHKGSEYVQASVYGGKFLYFLNKKHGWALP 1558
Qy 1560 FLIQGVSVSYGHIKHDTTFLYPSIHERNKGWEDLGWLADLRISMDLKEPSKDSKRITVY 1619
Db 1559 FLIQGVSVSYGHIKHDTTFLYPSIHERNKGWEDLGWLADLRISMDLKEPSKDSKRITVY 1618
Qy 1620 GELEYSSIRQKQFTIEDYDPRHFDCCAYRNLSLPGCAVEGAIMNCNITLMYKALAYMP 1679
Db 1619 GELEYSSIRQKQFTIEDYDPRHFDCCAYRNLSLPGCAVEGAIMNCNITLMYKALAYMP 1678
Qy 1680 SIYRNPVCKYRVLSNNEAGOVICGVPTRTSARAESYTQLYLGPFWTLTYGNVTDVGNVT 1739
Db 1679 SIYRNPVCKYRVLSNNEAGOVICGVPTRTSARAESYTQLYLGPFWTLTYGNVTDVGNVT 1738
Qy 1740 LSQMTSCGARMIF 1752
Db 1739 LSQMTSCGARMIF 1751
RESULT 2
Q9PJY2 PRELIMINARY; PRT; 1672 AA.
AC Q9PJY2:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN B/C FAMILY.
GN TC0694.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG; PubMed=10684935;
RX MEDLINE=20150255;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Ewison M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002338; AAF39510.1; -.
DR TIGR; TC0694; -.
DR InterPro; IPR003357; -.
DR Pfam; PF02385; OMP; 1.
SQ SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;
```

Query Match 72.6%; Score 6488.5; DB 2; Length 1672;  
Best Local Similarity 73.0%; Pred. No. 0;

Matches 1292; Conservative 145; Mismatches 214; Indels 119; Gaps 20;

```
QY 1 MKWLSATAFAAALPVSVCFCPEPEKELNFSRVETSSSTFTTETIGEAGAEYIVSGNASF 60
Db 1 MKWLSATAFAAALPVSVCFCPEPEKELNFSRVETSSSTFTTETIGEAGAEYIVSGNASF 60
QY 4 MKWLSATAFAAALPVSVCFCPEPEKELNFSRVETSSSTFTTETIGEAGAEYIVSGNASF 63
Db 4 MKWLSATAFAAALPVSVCFCPEPEKELNFSRVETSSSTFTTETIGEAGAEYIVSGNASF 63
QY 61 TKFTNI-----PTTD---TTTPNNSNSSSGE-----TASVSDSDSTTTTPDPKGGGAFY 109
Db 61 TKFTNI-----PTTD---TTTPNNSNSSSGE-----TASVSDSDSTTTTPDPKGGGAFY 109
QY 64 TNFTNPVKKPTTDDSTSTPTTSSAVDPTKIVIRASSSPNSGSDTSATPDPKGGGAFY 123
Db 64 TNFTNPVKKPTTDDSTSTPTTSSAVDPTKIVIRASSSPNSGSDTSATPDPKGGGAFY 123
QY 110 NAHSGVLSFWMTSGTSGSLTSLBIKMTGGGALFSGOGLLFTDLTSLITQNNLSQLSGGA 169
Db 110 NAHSGVLSFWMTSGTSGSLTSLBIKMTGGGALFSGOGLLFTDLTSLITQNNLSQLSGGA 169
QY 124 NEHSGILSPWARSVGSGLTSLNIKMTGGGALFSGOGLLFTDLTGLTQCNLSQLSGGG 183
Db 124 NEHSGILSPWARSVGSGLTSLNIKMTGGGALFSGOGLLFTDLTGLTQCNLSQLSGGG 183
QY 170 IFGGSTISLGIKATFSCNSAEVPAVPKPTPKAQTASSETSGSSSSGNDVSPSSPS 229
Db 170 IFGGSTISLGIKATFSCNSAEVPAVPKPTPKAQTASSETSGSSSSGNDVSPSSPS 229
QY 184 IFGGSTISFGINQATFSSNTAEV---VPEETPNPNPGTQTTSQ-----PSPTS- 231
Db 184 IFGGSTISFGINQATFSSNTAEV---VPEETPNPNPGTQTTSQ-----PSPTS- 231
QY 230 RAEPAAANLQSHFICATAPAAQTDDETSTPSHKPGSGGAIYAKGDLTTADSOEVLFSIN 289
Db 230 RAEPAAANLQSHFICATAPAAQTDDETSTPSHKPGSGGAIYAKGDLTTADSOEVLFSIN 289
QY 232 -----KVQSLFTYSSSTQANGSADQSTPSHKPGSGGAIYATGDLTISDSQEIYFSVN 284
Db 232 -----KVQSLFTYSSSTQANGSADQSTPSHKPGSGGAIYATGDLTISDSQEIYFSVN 284
QY 290 KATKDGALFAEKDVSFENITSLKVOTNGAEKGGAIYAKGDLISIOSKOSLFSNYSKQ 349
Db 290 KATKDGALFAEKDVSFENITSLKVOTNGAEKGGAIYAKGDLISIOSKOSLFSNYSKQ 349
QY 285 KASKDGGALFAEKWSEFNITLKVQNGAEKGGIYASGDLISIOSKOSLFSNYSKQ 344
Db 285 KASKDGGALFAEKWSEFNITLKVQNGAEKGGIYASGDLISIOSKOSLFSNYSKQ 344
QY 350 GGGALYVEGGINFQDLLEIRIKYNKAGTETKRTITLPSLKAQASAGNADAWASSPSQSGS 409
Db 350 GGGALYVEGGINFQDLLEIRIKYNKAGTETKRTITLPSLKAQASAGNADAWASSPSQSGS 409
QY 345 GGGALYIEGNVDFKOLEIRIKYNKSGTETKRTITLPSLKAQASAGNADAWASSPSQSGS 394
Db 345 GGGALYIEGNVDFKOLEIRIKYNKSGTETKRTITLPSLKAQASAGNADAWASSPSQSGS 394
QY 410 GATTVSDSGSSGSDSDETSVETPVVTAAGGGLYTDKNLSITNITGIEIANKKATDVGGG 469
Db 410 GATTVSDSGSSGSDSDETSVETPVVTAAGGGLYTDKNLSITNITGIEIANKKATDVGGG 469
QY 395 -----ASOSGPNNTPTPTPPVTAAGGGLYTDKNLSITNITGIEIANKKATDVGGG 445
Db 395 -----ASOSGPNNTPTPTPPVTAAGGGLYTDKNLSITNITGIEIANKKATDVGGG 445
QY 470 AYVKGTLTENSRLQFLKNSDKOGGGLYGEDNITLSNLTKTFOENTAKEEGGLFI 529
Db 470 AYVKGTLTENSRLQFLKNSDKOGGGLYGEDNITLSNLTKTFOENTAKEEGGLFI 529
QY 446 AYVKGTLTCKSHRLQFQKNSDKOGGGLYTEDITLNLTKTFOENTAKEEGGLYI 505
Db 446 AYVKGTLTCKSHRLQFQKNSDKOGGGLYTEDITLNLTKTFOENTAKEEGGLYI 505
QY 530 KGTDKALMTGLDLSFCLINNTSEKHHGGAFVTKETISQTVTSDVETIPGTPVHGETVITG 589
Db 530 KGTDKALMTGLDLSFCLINNTSEKHHGGAFVTKETISQTVTSDVETIPGTPVHGETVITG 589
QY 506 QGDDKTLTWTGLDLSFCLIDNTSATHGGGAVTKETISQTVTSDVETIPGTPVHGETVITG 565
Db 506 QGDDKTLTWTGLDLSFCLIDNTSATHGGGAVTKETISQTVTSDVETIPGTPVHGETVITG 565
QY 590 NKSTGGNGGVCTKRLALNSLISISGNSAABNGGAHTCPDPSFTADTAEOPAASAA 649
Db 590 NKSTGGNGGVCTKRLALNSLISISGNSAABNGGAHTCPDPSFTADTAEOPAASAA 649
QY 566 NKATGGSGGVCTKHLVLSNLQITISSEFASENGGAHTCPDPSFTADTAEOPAASAA 611
Db 566 NKATGGSGGVCTKHLVLSNLQITISSEFASENGGAHTCPDPSFTADTAEOPAASAA 611
QY 650 TSTPKAPYSTALSTPSSSTVSSLLTAAASSQASPATSNKETQDPNADTDLIDYVVDIT 709
Db 650 TSTPKAPYSTALSTPSSSTVSSLLTAAASSQASPATSNKETQDPNADTDLIDYVVDIT 709
QY 612 -----APTA---STPSTN-----QTAAPKDDKDFLIDYVVDIT 641
Db 612 -----APTA---STPSTN-----QTAAPKDDKDFLIDYVVDIT 641
QY 710 ISKNTA-KRGGGIYAKKAKMSRIDQLNISSENSATEIGGGICCKESLELDALVSLVTENL 768
Db 710 ISKNTA-KRGGGIYAKKAKMSRIDQLNISSENSATEIGGGICCKESLELDALVSLVTENL 768
QY 642 IDKNKATKKGAGYAKKALSRIDELNISDAAQETGGGFCCTESLELDTIASLVTKNL 701
Db 642 IDKNKATKKGAGYAKKALSRIDELNISDAAQETGGGFCCTESLELDTIASLVTKNL 701
QY 769 VKKEGGLHAKVTNINSLKSGFSFNKNSNSTGVATTASAPAAAALQAAAAAPS 828
Db 769 VKKEGGLHAKVTNINSLKSGFSFNKNSNSTGVATTASAPAAAALQAAAAAPS 828
QY 702 AKKEGGLHAKVTNINSLKSGFSFNKNSNSTGVATTASAPAAAALQAAAAAPS 761
Db 702 AKKEGGLHAKVTNINSLKSGFSFNKNSNSTGVATTASAPAAAALQAAAAAPS 761
QY 829 PA---TPTYSGVGGAIYGEKVTFSQCSGTCQFSGNOAIDNPNSSSLNVQGGAIYAKT 884
Db 829 PA---TPTYSGVGGAIYGEKVTFSQCSGTCQFSGNOAIDNPNSSSLNVQGGAIYAKT 884
QY 762 PAPAQTPTPYAGVGGAIYGEKVTFSQCSGTCQFSGNOAIDNPNSSSLNVQGGAIYAKT 821
Db 762 PAPAQTPTPYAGVGGAIYGEKVTFSQCSGTCQFSGNOAIDNPNSSSLNVQGGAIYAKT 821
QY 885 SLISGSDAGTSVIFSGNSVSTGKSTQTOIAGGAIYSPVTILNCPATFSNNTASTATPK 944
Db 885 SLISGSDAGTSVIFSGNSVSTGKSTQTOIAGGAIYSPVTILNCPATFSNNTASTATPK 944
QY 822 SLISAEADPTSVFSGNSVSTGKSTQTOIAGGAIYSPVTILNCPATFSNNTASTATPK 879
Db 822 SLISAEADPTSVFSGNSVSTGKSTQTOIAGGAIYSPVTILNCPATFSNNTASTATPK 879
QY 945 TSSEDGSSGNSIKDTIGGAAGTATLGSVSRFSRGNTADLGAAGTLANATPSATSGSO 1004
Db 945 TSSEDGSSGNSIKDTIGGAAGTATLGSVSRFSRGNTADLGAAGTLANATPSATSGSO 1004
QY 880 ---TNPPSGTSPKDTIGGAAGTATLGSVSRFSRGNTADLGAAGTLANATPSATSGSO 928
Db 880 ---TNPPSGTSPKDTIGGAAGTATLGSVSRFSRGNTADLGAAGTLANATPSATSGSO 928
QY 1005 NSITEKITLNGSFTFERQNAKRGAIYSPVSIKGNNTFNQNTSHDGSALYFTKDAT 1064
Db 1005 NSITEKITLNGSFTFERQNAKRGAIYSPVSIKGNNTFNQNTSHDGSALYFTKDAT 1064
QY 929 SNLTKITLNGSFTFERQNAKRGAIYSPVSIKGNNTFNQNTSHDGSALYFTKDAT 988
Db 929 SNLTKITLNGSFTFERQNAKRGAIYSPVSIKGNNTFNQNTSHDGSALYFTKDAT 988
```

```
RESULT 3
084419 PRELIMINARY; PRT; 1770 AA.
ID 084419;
AC 084419;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN C.
GN PMPC.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/OW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL; AB001315; AAC68011.1; -.
```



RESULT 4  
Q9PJY1 ID Q9PJY1 PRELIMINARY: PRT: 1460 AA.  
AC Q9PJY1  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE POLYMORPHIC MEMBRANE PROTEIN B/C FAMILY.  
GN TC0695.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_Taxid=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MOPN / NIGG.  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weldon J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL; AE002338; AAF39511.1; -;  
DR TIGR; TC0695; -;  
DR InterPro; IPR003357; -;  
DR Pfam; PF02385; OMP; 1.  
SQ SEQUENCE 1460 AA; 154781 MW; C1F033E7907AC3AC CRC64;

Query Match 35.8%; Score 3205; DB 2; Length 1460;  
Best Local Similarity 39.3%; Pred. No. 1.5e-162;  
Matches 741; Conservative 218; Mismatches 366; Indels 562; Gaps 33;

QY 1 MKWLSATVFAAALPVSFGFCPP-----EPKELNFSRVSTSTFTTETGEAGAAYVSG 56  
DB 1 MKFLSATVFAAALPVSITSASSVESQIETKDLNSSRTGSSSSQSFTEIIPENGAAYRVSG 60  
QY 57 NASFTKFTNIP-----TDTTTPNS 77  
DB 61 DVSFDSFNIPAEATLAISHKEOPNNEVVLSEENHQAFQDSQAQNOTENASGNSPSE 120  
QY 78 NSS-SSSGETAS-----YSDSDSTTTTP----- 100  
DB 121 NTNQSTTTETSTTDEQVQNDNESASVPTTETATAMRLPSYHLQTESLVEGATEEQ 180  
QY 101 -----DPKGGGAFYNAHSGVLSEFTRSGTGLSEIKMTGEGGAFISQGELEFIDL 153  
DB 181 NQPNQNTSSGGGAFYNSQGGPLSFINDPKDSSLTKIRVIGEGGATYSGPLSITGL 240  
QY 154 TSLITONLSQLSGGAFGGSTISLSGITKATFSCNSAEVPAPVKK--PTEPKAQTASET 211  
DB 241 KKLAKENLSQKAGGAICAESTISISSVDSIIFSKNTVTPPAANKPELPNDPSGNGNDG 300  
QY 212 SGSSSSSGND-----SVSPSSSRAEFAAANLQSHFICATATPA 250  
DB 301 SDDSNSGNTDSNESPNNSASNTGSENELSSSTPSAQLPNPA-----TPFLSSVSTNS 355  
QY 251 AQTDTETSPSH-KPGSGGAIYAKGLDITADSQEVLFSINKATKDGGAIFAEKDVSEFNI 309  
DB 356 QPIDTEPENAWHAESGGGAIYKSGKLSASSKEVYFDHNSATKNGGAIFGEEIALEKI 415  
QY 310 TSLKVOTNGAEKGGAIYAKGDLISQSSKQSLFNSYSKQGGGALVEGGINFQDLEIR 369  
DB 416 ASLKDSNTTGEKGAIAK----- 435  
QY 370 IKYNKAGTFETKKITLPSLKAQASAGNADAWASSPQSGSGATTVSDSGSSGSDSTS 429  
DB 436 ----- 435  
QY 430 ETVVPYAKGGGLYTDKNLSITNITGIIETANNKATDVGGAYVKGTLTENSHRLOFLKN 489

DB 436 ----- 435  
QY 490 SSDKGGGIYGEDNITLSNLTGKTLFQENTAKEEGGLFIKGTDKALTWGLDSFCLINN 549  
DB 436 -----TVTSLDIKNTLIFVNTAK----- 454  
QY 550 TSKHGGGAFVTKETISQTSVTVETIPTPVHGETVITGNKSTGGNGGCVCTKRLALS 609  
DB 455 -----TPEE-----NS 460  
QY 610 LQSIISGNSAENGGAHFCPSFPTADTAQPAASAATSTPKAPVSTALSTPSST 669  
DB 461 LKSSQLNNQNPSEE-----EHQDTSEGESQSLETS-----PITQDS--ASSH 502  
QY 670 VSSLTLAASSQASPATSNKETODPNADTDLIDYVVDTTISKNATAKGGGIYAKKAKMS 729  
DB 503 VAIFRISIASS-----SQSNSE-NIPNAD----- 525  
QY 730 RIDQLNISENSATEIGGICCKRESLELDALVLSVTENLVKEGGLHAKTVNISLKSG 789  
DB 526 -----GSTSAGG----- 532  
QY 790 FSPSNKANSSTGCVATTASAPAAAAAASLQAAAAAASPSPATPTYSGV--GGAIYGEKV 847  
DB 533 -----DAGSSQPSPTPGSDSSINHVGGAIIYGEAV 563  
QY 848 TFSQCSGTQCFSGNOAIDNNPSSSLNVOGGAIYAKTSLSIGSSDAGTSYIFSGNSVSTG 907  
DB 564 KIENLSGYGTFSNNNAVDHQISGTSVDLGGAIYAKTSLTIDSGNSGTTTFSEN--TTS 621  
QY 908 KSOTGQIAGGAIYSPVTLNCNPATPSNNTASIACTPKTSESDGSSGNSIKDTIGGAIF--AG 966  
DB 622 SKETTGQVAGGAIYFSPVITTPVTFKNSAINAT--TSSK-----KDTFGAIGAI 671  
QY 967 TAITLSCVRFSCNTADLGAACIGCTLANANPSPATSGSONSITEKIENLNGSFIFERNQAN 1026  
DB 672 STVLSKGFARFSENIADLGAISGLVPTQD-----AETVQLTTSYGFERNKAL 720  
QY 1027 KRCAIYSPSVSIKGNITFNQNTSTHDGSAIYFTKDATIESLGSVLFTGNVNTATQASSA 1086  
DB 721 KRATVAPIVSIKAHATFDQNISAEEGSAIYFTKEATIESLGSVLFTGLNLTVPQISTTV 780  
QY 1087 -TSGQNTANYGAAIFGDPGTTQSSQTDAL-LTLASSGNITFSNNSLQNNQDTPASK 1144  
DB 781 LTSG---NTSKYGAALFGQIANASGSDTNLPLKLIIASGNIISFRNNEYRDPATNTQST 837  
QY 1145 FCSIAGYVKLSLQAAKGKTSFPDCVHTSTKKTGSTQNVYETLIDINKEENS----NFYTC 1200  
DB 838 FCSIAGDIKLTMOAEGKVISFDDAIRTSTKGTGLASAYDTLIDINKSNDGSIASFTG 837  
QY 1201 TIVFSELHENKSYIPQNAITLHNGTLVLKTELHVYFPEQKSGSLIMEPGAIVLSNQNI 1260  
DB 898 TINFSELHENKSYIPQNVVLHSGSLILKANTELHLVSPDQKSGSLIMEPGSVLSNQDI 957  
QY 1261 ANGALAINGLTIDLSMGTQPAQ--EIFSPPELRIVATSSASGSGSVSSSIPTNPKRIS 1318  
DB 958 ADGSLVNVSLTIDLSVGRNSAGDNIEMPELRIVDTSTN---SGNSSSTPSSNTTP 1013  
QY 1319 AAVPSSA-----ATTPTMSENKVFITGDLTIDPNGNFYQNPMLGSLDVL 1366  
DB 1014 NSTPTAQAPISKNAFAATTTPTTPTTGNIVFLNGVIKLDIPNGTFFQNPALGSDQKISL 1073  
QY 1367 IKLPTNTSDVQVYDLTSLGDLFPQKGMGTWLTDSNP-OTGKLAQRTPTTYRRVYIPR 1425  
DB 1074 LVLPSTQTKLAQKVVLTGDISPKKYTGTLTLDPOOLQNGVIOALWTFKSYQWYIPR 1133  
QY 1426 DNHFYANSILGSONSMIVVKQGLINNMNAREDDIAYNNFWYSGVGTFLAQOQTPLSEE 1485  
DB 1134 DNHFYANSILGSONSMATVKQGLINDKLNARFEDEVAYNNLWISGLTMLSQRGQRSEE 1193  
QY 1486 FSYYSRGTSVAIDAKPRQPFILGAASFVKYKTKAKIKMHNYPHKSEYYSQASVYGGK 1545  
DB 1194 MTYYSRGASVALDKPTQDLIIIGAASFVKYKTKAKIKMHNYPHKSEYYSQASVYGGSP 1253



Db 1205 -TGTLFSELHENKSIYIPQKVTFAHGNILILGNLAELSUVVSFTQSPGTTITMGPGSVLSN 1263  
QY 1258 QNIANGALAINGITLIDLSMGTPQAGEIFESPPELRIVATTSASGSGVSSSIPTNPKRI 1317  
Db 1264 HSKAAGGAINVVIDFSEIVPTKONATVAPPTLKLVSTNAD-----1306  
QY 1318 SAAPVSGSAATPTMTSENKVFITGDLTLIDPNCNFYQNMPLGSDLDVPLIKLPTNTSD-V 1376  
Db 1307 -----SKDKIDITGVTLLDPNGLYQNSYLGEDRDITLFDNDSAGAV 1351  
QY 1377 QYVDLTLSGLDLPFGKGYMCTWTLDSPQTKLQARWTFDTYRWVYIPRONHYANSILG 1436  
Db 1352 TATNVTLQNLGAKKGYLTWLNPNSSGSKILKWTFDKYLWRVPYIPRDNHFIINSWG 1411  
QY 1437 SONSIMVYKQGLINMNNAREDDIAYNFWVSGVGTFLAQOQTPLSSEFYSYRGTSTA 1496  
Db 1412 AQNSLVYKQGLINMNNAREDDAFNFWASAGSFLRKEVSRNSDSFTVHGRYTA 1471  
QY 1497 IDAKPRQDILGAARSKIVGKYATKMKMNYFHKSGEYQASVYGGKFLYLLNKQHW 1556  
Db 1472 VDAKPRQEPILGAARFQVFGHAESYVHLNRYKHKGSGHSTQASLYAGNIFYFPAIRSR-- 1529  
QY 1557 ALPELQGVVYGHKIDHTTLYPSYTHERNKGWEDLGLADLRISMOLKEPSKSSKRI 1616  
Db 1530 --PILFQGVATYGYWQHDTFTTYPSIEEKNMANWDSIAWFLDLRFSDVLEKEPOPHSTARL 1587  
QY 1617 TVYGELEYSSIRQKOPTEDIDPRHFDCCAYRNLSLPVGCAGEAIGMNCNIIYMKLALA 1676  
Db 1588 TTYTAEYTRIQERFELDYDPRFSACSIGNLAIPTGFSVDGALAWREIILYKNVSA 1647  
QY 1677 YNPSYRNPNVCKYRVLSNEAGQVIGCVPTRTSARAERYSTOLYLGPFWTLYGNYTIDVG 1736  
Db 1648 YLPVLIRNPNKATYEVLSRKEKGNVYVLPTRNAAREVSSQIYLGSYWTLYGTYTIDAS 1707  
QY 1737 MYTLOMTSCGARMIF 1752  
Db 1708 MNTLVOMANGIRFVF 1723

RESULT 6  
Q9RB59 PRELIMINARY; PRT: 1723 AA.  
ID Q9RB59  
AC Q9RB59  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE POLYMORPHIC MEMBRANE PROTEIN B FAMILY.  
GN PMP\_20.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,  
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,  
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,  
RA Ishii K., Shiba T., Hattori M., Kuhara S.,  
RT "Comparison of outer membrane protein genes omp and pmp in the whole  
RT genome sequences of Chlamydia pneumoniae isolates from Japan and US";  
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
DR EMBL; AB033817; BAA85968.1; -.  
DR EMBL; AP002547; BAA98746.1; -.  
DR InterPro; IPR001254; -.  
DR MSRIDQINISENATEIGGICCKESLELDALVSLSVTENLVKGGGLHAKTVNISHLK 787

DR InterPro; IPR003357; -.  
DR InterPro; IPR003368; -.  
DR Pfam; PF02385; OMP; 1.  
DR Pfam; PF02415; DUF145; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
SQ SEQUENCE 1723 AA; 179593 MW; 4507D5B8A0E3EF01 CRC64;  
  
Query Match 34.8%; Score 3111.5; DB 2; Length 1723;  
Best Local Similarity 37.6%; Pred. No. 1.8e-157;  
Matches 727; Conservative 284; Mismatches 528; Indels 397; Gaps 45;  
  
QY 1 MKWLSATAVFAAALPVSVSGFCFPEPKELNFSRVETSSSTT-----FTETIGEA-GAEYI 53  
Db 1 MKWLPAVFAAALPVSVSGFCFPEPKELNFSRVETSSSTT-----FTETIGEA-GAEYI 60  
QY 54 VSGNASFTFTNPTTDTTTPINSNS-----SSSGET-----86  
Db 61 IVGDITFTFTNIPVPVPTPDANDSSSSKGGSSSGATSLIRSNLHSDFDFTKSVL 120  
QY 87 -----ASVSDSDSTTTTTPDPKGGGAFY-NAHSG 114  
Db 121 DLYHLFPSPASNTLNPALLSSSSGSSSSSSSSGASAVVAADPKGGAAYFSEANG 180  
QY 115 VLSFMTSRGTEGLTSEIKMTGEGGAIFSQGELLEFLDTLTIONNLSQSGAIF---171  
Db 181 TLTFTTDSGNPGSLTLQNLKMTGDGAAYSKGPLVFTGLKLNLTFTGNEKSGGAAAYTEG 240  
QY 172 -----GG-----STIS 177  
Db 241 ALTQAIQEAIVTFTGNTSAGQGGAIYVKEATLNFALDSLKFENKTSQAGGGIYTESTLT 300  
QY 178 LSGITKA-TFSCNSARVPAPVKKPTKPAQTASETSGSSSSGNDSDSVSPSSSRAPFAA 236  
Db 301 ISNIKASIEFISNKASVPAPAPEPTSP-----APSSLINSTTIDTTLQTRAASATPAVA 355  
QY 237 NLOSHFICATAPAAQTDTTETSTPSHKPGSGGAIYAKGDLTI-----278  
Db 356 PV-----AAVTPPTISTQETA-----GNGGAIYAKQGISISTFKDLTFKSNASVDAT 403  
QY 279 -----ADQEV-----LFSINKATDGGAIFAEKDVSPENITSLKV 314  
Db 404 LTVDSSTIGESGGAIFAADSIQOCTGTTLFSGNTANKSGGGIYAVGQVTLIEDIANLKM 463  
QY 315 QNGAEEKGAIYAKGDLSTQSSK-QSLFNSNYKOGGAIYVEGGIFQDLDEIRIKYN 373  
Db 464 TNNCKEGEGAIYTKKALINNGAILTFTSGNTSTONGGAIFAVGGITLSDLVEVRFKN 523  
QY 374 KAGTFETKKTITLPSLKAQASAGNADAWASSSQSGSGATTVSDSGSSGSDSDTSETVP 433  
Db 524 KTCNY-----SAPITKA---ASNTAPVWSSSTTAASPAV-----PAAAAAP 561  
QY 434 VT--AKGGGLYTDKNLSITNITGIIAIANNKATDVGCGGAYVKGTLTNCENSHRLOFLKNS 491  
Db 562 VTNAAKGGLYSTEGLTVSGITSLFENNECONQGGIYAVTFTQCSDSHRLOFTSNKA 621  
QY 492 DXOGGIIYGEDNITLNLTKTLFOENTAKEEGGFIKGTDKALMTGLDSCFLNNTS 551  
Db 622 ADEGGGLYCGDDVTLNLTKTLFOENSEKGGGLSL-ASGSLTWTLSLESCFLNANTA 680  
QY 552 EKHGGGAFVYKEISQTYT---SDVETIPGTPVHGETVITGNKSTGGNGGVCVKRLALS 608  
Db 681 KENGGGANVPENIVLTFTPTTNEPAPVQOPVYGEALVTGNTAT-KSGGGIYTKNAAFS 739  
QY 609 NLOISISGNSAENGCGAHTCPDPTADTADQPAASAASTTPKSPAVSTALSTPSS 668  
Db 740 NLSSVTFDQNTSSENG-----ALLTQKAAD-----765  
QY 669 TVSSLLTLLAASSQASPATSKNETQDPNADTDLIDYVVDTTISKNTAK-KGGGIYAKKAK 727  
Db 766 -----KDCSFTYITNVTNNTATGNGGIAGGAKAH 797  
QY 728 MSRIDQINISENATEIGGICCKESLELDALVSLSVTENLVKGGGLHAKTVNISHLK 787

```
Db 798 FDRIDNLTVQSNQAKK-CGGVLEDALEKIVITGVSQNTATESGGGIYAKDIQALP 856
Qy 798 SGFSFNKANKSSSTGVATTASAPAAAAASLQAAAAAPSPPTTYSGVVGAIYGE-K 846
Db 857 GSFTITDNKTSLT-----TSNLYGGGIYSSGA 886
Qy 847 VTFSCSGTCFSGNQADNNPSSSLNVQGGAIYAKTSLSIGSSDAGTSYFSSNST 906
Db 887 VTLNISTGFTIGTNSVINTATSQDA-DIQGGIYATLSLI--NQCWTFILFSSNSAAT 943
Qy 907 GKSQTTQGIAGGIYSPVTVL---NCPATFNSNTASTATPKTSSDSSGNSIKDTIGGA 963
Db 944 KKTSTTKIAGGAIFSAVAVTIENNSOPIFLNNSAK-----SEATTAATAGN-KDSCGA 997
Qy 964 IAGTAITLSGVSRR--FSNTADLGAAG--TLANANTPSATSSQNSITEKITLNGSFI 1019
Db 998 IAAANSVTLTNNPETTFKGNAYETGAIGCIDLTNGSPPRKYS-----IADNGSVL 1047
Qy 1020 FERNOA-NKRGAIYSPSVISKGNITFQNTSTHDGSAIYFTKDATIESLGSVLFTGNV 1078
Db 1048 FQDNSALNRGAIGYETIDISRTGATFIGNSKKHGSAICCTALTALPNSQLIFENKV 1107
Qy 1079 TATQASSATSGONTNANYGAIFGDPQTQSSODAILTLASSGNITTFNNSLQNNQ 1138
Db 1108 TETTATTAKS-----INNLAGAIYGNNET-----SDITISLSAENGSIFFKNLNC----- 1152
Qy 1139 DTPASKFCSIAGYVKL-SLOAKGKTIISFFDCVHTSTKTKTSTQNVYETLIDNKEENSP 1197
Db 1153 -TATNKYCSIAGNKFTFAIESAGAKAISFYDAVNVSTKETNA-----QELKLNKATYS-- 1204
Qy 1198 YGTIVFSEHKNKSIYIPQNAIHLNGFLVLKEKTELHVVSFEQKSGKLIMEPGAVLSN 1257
Db 1205 -TGILFSGELHENKSIYIPQVTAHGNLILGKNAELSVSFTQSPGTTIMPGSVLSN 1263
Qy 1258 QNIANGALAINGLTIDLSMTGTQAGETSPPELRIVATTSASGGSGVSSSIPNPKRI 1317
Db 1264 HSKAAGGIAINNVIIDEFSEIYPTKDNATVAPPTLKLVSRTNAD----- 1306
Qy 1318 SAAVPSGSAATPTMSENKVFELTGDLTLDIPNGFYQNPMLGSLDVLPLKPTNTSD-V 1376
Db 1307 -----SKDKIDITGTVTLIDPNGNLYQNSYLGEDRDITLFDNIDSASGAV 1351
Qy 1377 QVYDITLGSGLDPQKGYMGWTLDNSNPOTGKLOARWTFDYRRWYIIPRONHVFANSTLG 1436
Db 1352 TATNVTTLQGLNKAAGYLGTLNWDNPSGSKIILKWTFDKYLRWPIIPRONHFIINSIW 1411
Qy 1437 SONSMIVVQGLINMLNARFDDIAYNFWVSGVGTFLAQQGTPLSEFSYYSRGTSVA 1496
Db 1412 AONSLVTYKQILGNMLNARFEDPAFNFWASAIKSPFLRKEVSRNSDSFTYHGRGYTAA 1471
Qy 1497 IDAKPRODFILGAAFSKIYGVKTKAIKMHNYFHKGSYYSQASVYGGKFLYELLNKHQHW 1556
Db 1472 VDAKPRQFELGAAFSQVGHAESEYHLDNTKHKGSHTQASLYAGNIYFFPAIRSR-- 1529
Qy 1557 ALPELIQVWSYGHIKHDTTLYPSIHERNKGDWEDLGWADLRISMOLDKEPSKDSKKRI 1616
Db 1530 --PILFQGVATYGYMQHDTTYYPSIEKKNANWDSIANWFLDRFSVDLKEPQPHSTARL 1587
Qy 1617 TVYGELEYSSIRQKQFTIEDYDPRHFDCCAVRNLSLPGVCAVEGAIMCNILMYNKLALA 1676
Db 1598 TIFYTEAEYTRIQEKFTELDYDPRFSACSYNLAIPTFGVSVDGALAWREILYLNKVSAA 1647
Qy 1677 YMPSTYRNPNVCKYVLSSNAGOVICQVPTRTSARAEYSTQLYLGPWTLGYNTIDVG 1736
Db 1648 YLPVILRNPNATYEVLTSEKKGNNVNVLPTRNARAEEVSSQIYLGSTWTLGYNTIDAS 1707
Qy 1737 MYTLQMTSCGARMIF 1752
Db 1708 MNTLVQMANGGIRFVF 1723
```

RESULT 7

```
Q9K2C1
ID Q9K2C1 PRELIMINARY; PRT; 1732 AA.
AC Q9K2C1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN B/C FAMILY.
GN CP0212.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eissen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002181; AAF38082.1; -.
DR TIGR; CP0212; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR003357; -.
DR Pfam; PF02385; OMP; 1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
SQ SEQUENCE 1732 AA; 180631 MW; DB02D918C777B525 CRC64;
```

Query Match 34.8%; Score 3111.5; DB 2; Length 1732;

Best Local Similarity 37.6%; Pred. No. 1.8e-157; Indels 397; Gaps 45;

Matches 727; Conservative 284; Mismatches 528;

```
Qy 1 MKWLSATAFAAALVPSVSGFCFPPKELNFRVETSSSTT-----FTETICEA-GAEYI 53
Db 10 MKWLPATAFAAALVPSVSGFCFPPKELNFRVETSSSTT-----FTETICEA-GAEYI 53
Qy 54 VSGNASFTKFTNIPTDTTPTNSNS-----SSSGSET----- 86
Db 70 IVGDITFTFTNIPVPVVTDPANDSSSSSKGSSSGSGATSLIRSSNLHSDFDFTKDSVL 129
Qy 87 -----ASVSESDSTTTTDPKGGAFY-NAHSG 114
Db 130 DLYHLFFPSASNTLNPALLSSSSSGSSSSSSSSSSSSSASAVVAADPKGAFAFTNEANG 189
Qy 115 VLSFTMRSGTGLSLSEIKMTGEGGAIFSQCELLFTDLTSLITQNNLSQLSGGAIF--- 171
Db 190 TLTFTDTSNGPGLTLQNLKWTGDGAAIYSGPLVFTGLKNLFTTGNESQKSGGAAYTEG 249
Qy 172 -----GG-----STIS 177
Db 250 ALTTQAIWEAVTFTCNTSAGCGGAIYVKEATLFNALDSLKEFKNTSGOAGGIYTESTLT 309
Qy 178 LSGITKA--TSCNSAEVPAPYKVPKPEKQAOTASSETSGSSSSGNDSSVSSSRAEPAAA 236
Db 310 ISNTKSTIEFTSNKASVPAPAPEPTSP-----APSSLNSTIDTSTLTQTRAASATPAVA 364
Qy 237 NLSHFICATAPAAQTDSTSTPSHKPGSGGAIYAKGDLTI----- 278
Db 365 PV-----AAVTPPISTQETA-----GNGGAIYAKGISISTFKDLTFKNSASVDAT 412
Qy 279 -----ADSOEV-----LFSINKATKGGAIFAEDKVDSENTISLKV 314
Db 413 LTVDSSTIGESGAIFAADSIQIOCTGTTLFSGNTANKSGGIYAVGVQVLTEDIANLKM 472
Qy 315 QTNGAEKGGAIYAKGDLDSIOSK--QSLFNSNYKQGGGAIYVEGGINFQDEETIRYN 373
Db 473 TNNTCKGEGGAIYTKKALTINNGALLITFSGNTSTDNCGAIFAVGGITLSDLVVRFKSN 532
Qy 374 KAGTFETKKITLPSLKAQASAGNADAWASSSPQSGSGATTVSDSGDSSGSDSTETVP 433
```



Db 172 AVSAYSIDAKTTTAAALDQNTSTK-----NGGALCSTANTTVOGNSGTVTFFSSNTATDK 226  
QY 295 GGAIFA-EKDVSEFNITSLKVOTNGAEKGGAIYAKGDLISQSSKSLFNSNYSKQGGGA 353  
Db 227 GGGIYSKEKDSLDANTGVVTFKSTAKTGGAWSDDLALTGNTQVLFQEN-----278  
QY 354 LYVEGGINFQDLEERIKRYKAGTPEKTKITLPSLKAQASAGNADAWASSPQSGGA--411  
Db 279 -----KTTGSAQA-----NNPEGCGGAIC 298  
QY 412 ---TTVSDS---GSSSSGSDSTSTVPVYAKGGGLYTDKNLSITNITGIEIANKATD 465  
Db 299 CYLATATDKTGLAISQNOEMSFSTNT--TTANGGAIYATK-----336  
QY 466 VGGGAYVKGTLTCEHSHRLQFLKNSDKQ-GGGIYGE-DNITLSNLTGKTLFQENTAKEE 523  
Db 337 -----CTLDGNTTTLTFOONTATAGCGGAIYETEDFSLKSGTGTVTFTNTAK-T 385  
QY 524 GGGLFIKGTDKA-----LTMTGLDSFCLINNTSEKHGGGAFVTKETISQTYTSDVETIPGI 578  
Db 386 GGALYSKGNLSLTGNTNLLFSGNKATGSPNSANQEGCGGAILAFIDSGVS DK--TGL 442  
QY 579 TPVHGTEV-ITGNKSTGGNGGVCVKRLALSNLQISISGNSAENGGAHTCPDSFPFTA 637  
Db 443 STANQEVSLTSNAAT-VSGGAIYATKCTLTGNGSLTFDGNAGTSGGAIYETEDFTL- 500  
QY 638 DTAEQPAAASAATSTPKSAPVSTALSTPSSSTVSSITLLAASSQASPATSNKETODPNAD 697  
Db 501 -----TGSTGTV-----507  
QY 698 TDLIDYVVDVTTISKNTAKGGGIYAKKAKMSRIDQLNISENSATEIGGICCKESLELD 757  
Db 508 -----TFSTNTAKTGALYSK-----GNN-----526  
QY 758 ALVSLSVTENLVKGGGLHAKTVNISLKSFGFSNKNKANSSTGTGAVTASAPAAAAAS 817  
Db 527 ---SLSGNTNL-----FSGNKATGSPNSANQEGCGGAILSF 561  
QY 818 LQAAAAAPPSPATPTYSGVVGGAIYGEKVTFQCSCGTQCFPSQNOAIDNPNQSSINVOG 877  
Db 562 LESASVST-----KKGLWIENENVLSGNTA-----TVSG 592  
QY 878 GAIYAKTSLISGSDAGTSYIFSGNSVSTGKSQTTGQIAGGAIYSPPTVTLNC-----PAT 932  
Db 593 GAIYATKCALHCN---TTLTFDGN-----TAETAGGAIYETEDFTLTGSGTGT 639  
QY 933 FSNNTASTATPKTSEDGSSGNSIKDTTGGATAGTATILSGVSRFSGNTADLGAAGHLA 992  
Db 640 FSTNTA-----KTAGALHTKGN-----TSFTKNKALVFSGNSA-----TA 674  
QY 993 NANTPVSATSGQNSI-----TEKITL-ENGSTIFERNOANKR-GAIYSPSVSIKG 1040  
Db 675 TATTTDQEGCGGAILCNLISESDIATKSLTLENESLFINNTARSGGIIYAPKCVISG 734  
QY 1041 N-NITFNQNTSHDGSAIYFTKDKATIESLGSVLFTGNNVNTATQASSATSGQNTANTYGA 1099  
Db 735 SESINFQDNTAETSGGAIY-SKNLSITANGPVSTNN-----SGGKG-----GA 777  
QY 1100 AIFGDPGTQSQTDAILTLTLASSGNITFNSNLSQNNQGDTPASKFCSTAGVVKLSLQAA 1159  
Db 778 IYIADSGE-----LSLEAIDGDTFSGN--RATEGSTPNSIHLAGAKAITKLAAA 826  
QY 1160 KGKTSFFDCVHTSKTKSTQNVYETLDIN-----KEEN-----SN 1196  
Db 827 PGHTIYFYDPITMEAPASGTT---TEELVINPVKAIYVPPQPKNGPTASVPVVPVAPAN 883  
QY 1197 PYTGTVFSS-ELHENKSIYPQNAI-----LHNGTLVLKTEKTLHVVSFEQKESGLK 1247  
Db 884 PNTGTVFSSGKLPSQDASIPANTTTILNOKINLAGGNVVLKEGATLQVYSFTQPDSTV 943  
QY 1248 INEPGAVL--SNQNTANGALNGLTIDLSSMGTPQAGEIFSPPELRIVATTSASGSGG 1305  
Db 944 FMDAGTTLTTLTTNTDGSIDLKLSNLDAALDGKR-----MITIAVNSTSGGLK 993

QY 1306 VSSSIPTNPKRISAAVPSGSAATTTMTSENKVFLTGLDTLIDPNGNFYQNPMGLSDLDVP 1365  
Db 994 IS-----GDLKFHNNEGSFYDNPGLKANLNP 1020  
QY 1366 LIKLTPTNTSDVOVYDLT--LSGDLFPQKGYMGTTWLDNSNPQTG-----KLQARWTFDTYR 1418  
Db 1021 FIDLSTSGTGNLDDFNPTPSSMAAPDYQGSWTL--VPKVGAGGKVTLVAEQALGYT 1078  
QY 1419 -----RWYIIPRDNHIFYANSILGSONSMIVYKQGLNNLNNAREDDTAYNFWVSGVGT 1473  
Db 1079 KPPELRATLVP-----NSLWNAYVNIHSIQOETA-----TAMSDAPSHPGIWTGIGN 1126  
QY 1474 FLAQOCTPLSEEFYSYSGTSAIDAKPRODFILGAATSKIVGKTK--AIKKMHNVFHKG 1531  
Db 1127 AFHQDKQKENAGFRILSRGIYVGGSMTPPQETTFPAVASQLFKSKDYVVDIKSQVYAG 1186  
QY 1532 S---EYSQASVYGG---KFLYFLLNKHQHWALPFLIOGVVSYGHIKHDTTLYPSIHER 1585  
Db 1187 SLCAQSSYVIPLHSSLRHRVLSKVLPELPG-ETPLVLHGQVSYGRNHHNMTTKLAN-NTQ 1244  
QY 1586 NKGWEDLGLWADLRISMDLKEPSDKSKRITVYG---ELEYSSIROKQFTIDYDPRHF 1642  
Db 1245 GKSDWDHSFAVEVGSL---PVDLNYRYLTSYSPYVKLQVSVNKGQFQEVAAADPRIF 1300  
QY 1643 DDCAYRNLSLPGVCAV--EGALMNCNLMYNKALAYMPSIYRNPNPVCKYRVLVSNEAQ 1700  
Db 1301 DASHLVNVSIPMLGTLFKHESAKPPSALL---LTGAYDAYDRDHPHC-LTSLTNGTWS 1355  
QY 1701 VICGVPTRTSARAESTQLYLGPFWTLYGNVTIDVGMVTLTSLQMTSCGARMIF 1752  
Db 1356 TFATNLSHQAFPAEASGHLKLLHGLDCFASGSCELSSSSSYNANCGRYSF 1407

## RESULT 9

Q9JRW2 ID Q9JRW2 PRELIMINARY: PRT: 1276 AA.  
AC Q9JRW2:  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G/I FAMILY (POLYMORPHIC MEMBRANE  
DE PROTEIN G FAMILY).  
GN PMP\_6 OR CP0309.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA."  
RL Nucleic Acids Res. 28:2311-2314(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
Berry O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
Werry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,  
McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39."  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AP002546; BAA98652.1; --  
DR EMBL: AE002193; AAF38166.1; --  
DR TIGR: CP0309; --  
DR InterPro: IPR003357; --  
DR InterPro: IPR003368; --









```
Db 1144 MSFVSGDEASAEISNLSVSDQLQHVVTPLEEDTY---GHMGDMS-EAKIQDGTLVISW 1199
Qy 1413 TFDYRRWVYIPRDNHFFYANSILGSONSMIVV-----KQGLNNMLNNARFDDIAYN--- 1464
Db 1200 NPTGYR-----LDPOKAGALVFNALWEEGAVLSALAKNARF---AHNLTA 1240
Qy 1465 -----NFWVSGVGTFF--LAQCGTPLSEERSYYSRGTSVAIDAKPRQDFILGAAPSK 1513
Db 1241 QRMEDYSTNVMGFAGFGFTLSAENLVAIDGYKYGAGGASAGVDIQLMEDFVLGVSGAA 1300
Qy 1514 IVGKTKAIKKMHNHYFKHKGSEVSQASVYGGKFLYFLNKHQGWALPFLIOGVVSYGHKH 1573
Db 1301 FLGKMSQKDAEVSRAK-----VVGSVYTGFL-----AGSWFFKGQYSLGETQN 1345
Qy 1574 DTTTLTYSIHERNKGDWEDLGWLADLRIS-MDLKEPSKDS--SKRIIVYGELEYSSIROK 1630
Db 1346 DMKTRYGVLGE-SSASWTSRGVLADALVEYRSLVGPVRPTFYALHFNFPYEVSVASMKEP 1404
Qy 1631 QFTIEDYDPRHFDCCAYRNLSLPGCAVEGAIMNCNLTMYNKLALAYMPSIYRNPNVCKY 1690
Db 1405 GFTEQGREARSFEDASLTNITPLGMKFELAFIKGQFSEVNSLGSISYAWEAIRKVEGGAV 1464
Qy 1691 RVLSNNSAGVQICVGPRTSARAEYSQLYLGPPTWLYGNVTIDV---GMYT-----LS 1741
Db 1465 QLL---PAGFDWEGAP-MDLPRQELRVALENNTWWSYFTVLGLTAFCGGFTSTDSKLG 1520
Qy 1742 QMTSCGARMIF 1752
Db 1521 YEANTGLRLIF 1531

RESULT 13
Q9RB66 PRELIMINARY; PRT; 930 AA.
AC Q9RB66;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
  PROTEIN G FAMILY).
GN PMP_8 OR CP0307.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxId=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
  Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
  "Comparison of whole genome sequences of Chlamydia pneumoniae J138
  from Japan and CWL029 from USA.";
  Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
  Berry K., Bass S., Linner K., Weidman J., Khouri H., Craven B.,
  Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
  McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
  "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
  pneumoniae AR39.";
  Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AF002546; BAA98654.1; -;
DR EMBL; AE002193; AAF38164.1; -;
DR TIGR; CP0307; -;
DR InterPro; IPR003357; -;
DR InterPro; IPR003368; -;
DR Pfam; PF02385; OMP; 1;
DR Pfam; PF02415; DUF145; 1;
SQ SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;
```

```
Query Match 7.4%; Score 665.5; DB 2; Length 930;
Best Local Similarity 25.0%; Pred. No. 17e-27;
Matches 292; Conservative 150; Mismatches 389; Indels 335; Gaps 48;

Qy 658 VSTALSTPSSSTVSSLLTLLAASSQASPATS-----NKETQDPNADPDLIDYVVD 707
Db 10 ISSTLVTP---ILLSIATYCADASLPTDSPDAGGAGGTFTPKSTADANG-----TNYVLS 61
Qy 708 TTISKNTAKKGGIYAKKAKMSRIDQLNISENSATETGGGICCKESLELDALVLSVTEN 767
Db 62 GNVYINDAGK-----TAUTG-----FTET 83
Qy 768 LVKREGGLHAKTVNISNLKSGFSNNKANSSGTGVATTASAPAAAAAASLQAAAAAPS 827
Db 84 -----TGDLFTGKGYSFSTVDAGSN-----AGAAASTTADKALTFT 122
Qy 828 SPATPTVSGVVGGAIYGEKVTFSQC-----SCTCFSGNQAIIDNNPSSSLNVQGA 880
Db 123 GFSNLSTFAAPGTTVAGSKSTLSSAGALNLTDCNTILFSQNV-----NEANNNGAI 175
Qy 881 YAKTSLSIGSDAGTSYIFSGNSVSTGKSTQTTGOIAGGAIYSPVTLNCPATFNSNTASI 940
Db 176 TAKT-LSI-----SGNTSSITFTSNSAKKLGAIYS-----SAAASISGNTGQL 218
Qy 941 ATPKTSSEDDSGSNSIKDTIGGAI---AGTAITLGSVSRFSGNTADLGAAGTILANATP 997
Db 219 VFNNKGETG-----GGALGFASSTITONSSLFTSGNTA----- 253
Qy 998 SATSGSONSITEKITLLENGSFIFFERNOANKRGAIY-----SPSVSIKGN-NITFNQNTS 1050
Db 254 -----TDAAGKGAIYCEKGTETPTLTISGNKSLTFAENSS 289
Qy 1051 THDGSAYFTKDATIESLGSVLFTGNVNTATQASSATSGQNTNTANYCAAIFGPGTQTS 1110
Db 290 VTQGGAI-CAHGLDLSAAGPTLFSNNRC-----GNTAAGKGGAIAIADSGS--- 334
Qy 1111 SQTDAILLTLASSGNITFSNNLSNNOGDTFASKFCSTAGYVKL-SLQAAKGKTIISFDC 1169
Db 335 -----LLSANQGDITFLGNTLTSTAPTSTRNAIYLGSSAKITNLRAAQGGIYYPDP 388
Qy 1170 VHTSKTKGTQNVYETLIDINKEENSNP--YTGTFVFSSE---LHENK-----SYIQPN 1218
Db 389 IASNT--TGAS---DVLITINQPSNPLDYSGTIVFESGEKLSADEAKADNFTSILKQP 442
Qy 1219 AILLNGTILVKEKTELHVVSFEQKEGSKLINEPCAVLSNQNIANGALAINGLTIDLSMG 1278
Db 443 LALASGTALKGNVELDVNGFTQTEGSTLLMQPGTKLKAD---TEAISLTKLVVDLHAL- 498
Qy 1279 TPQAGEIFSPPELRIVATTSASGGSGVSSSIPTNPKRISAAVPSGSAATPTTMEHK-V 1337
Db 499 -----EGNKSVS-----IETAGAIKTI 515
Qy 1338 FLTGDLTLDIPNGNFYQNPMGLSDLDVPLIKLPTNTSDVQVY-DLTLSGDL---FPQKGY 1393
Db 516 TLTSPLVFQDSSGNFYESTHTINQAFQPLVYVFTAATAASDIYIDALLTSPQTPEHYGY 575
Qy 1394 MG-----TWTLDSDNPQTKLARWTFDIY-----RRWYIPRDNHFYAN-----SILG 1436
Db 576 QGHWEATWADTSTAKSGTM--TWVTTGYNPNPERRASVVP--DSLWASFTDIRTLOQIMT 631
Qy 1437 SONSMIVVVKOGLINNMLNNAFFDDIAYNNFVSCVGTFLAQOQGPLSEESYSRGSVVA 1496
Db 632 SQANSIYQQRGL-----WASGTANFFHKDKSGTQNAQFHKHSYGIVG 673
Qy 1497 IDAKPRQDFILGAAFSKIVGKTKAIKKMHNHYFKHKGSEYSYQASVY--GGKFLYLLAKQH 1554
Db 674 GSAEDFSENIFSVAFCQLFGKDKDLFIVENTSH-----NYLASLYLQHRAFLGLPWPSP 728
Qy 1555 GWA-----LPFLIQGVVSYGHIKHDTTTLTPSTHERNKGDWEDLGLWDLRLISMOLKEP 1608
Db 729 GSITDMLKDIPILILNAQLSYSTKNDMDTRVTSYPEA-QGSWTNNSGALELGGSLALYLP 787
```



DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).  
GN OMPs.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. phila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC/CWL-029/VR-1310;  
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;  
RT "Identification of two novel genes encoding outer membrane complex  
RT associated-surface-layer proteins in Chlamydia pneumoniae.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ001311; CAA04671.1; -  
DR InterPro; IPR003357; -  
DR InterPro; IPR003368; -  
DR Pfam; PF02385; OMP; 1.  
DR Pfam; PF02415; DUF145; 1.  
KW Signal.  
FT SIGNAL. 1 15  
FT CHAIN. 16 >914  
FT NON\_TER 914 914  
FT SEQUENCE 914 AA: 95603 MW: 8BF33BAB680FF5E3 CRC64;

Query Match 7.3%; Score 655.5; DB 2; Length 914;  
Best Local Similarity 25.5%; Pred. No. 5.6e-27;  
Matches 282; Conserved 146; Mismatches 433; Indels 247; Gaps 47;  
QY 670 VSSLTILAAASSQSPATSKNETQDPNADYDLDLDYVVDYVDTISKNTAKKGGIYAKKAKMS 729  
DB 9 VLSSLTACFTSCSTVFMAENIGPDSFD-----GSTNT-----GYTPKNTTT 53  
QY 730 RIDQLNISENATEIGGGICKESLELDALVLSVTEINLVGEGGLHAKTVNISMCLKSG 789  
DB 54 GIDYTLTGDTITLQNLGDSAAITKGCFSDFTESLSFA-----GKGYLSFLNIKSSAESG 106  
QY 790 FSPS-NKANSSTGVATTASAPAAAASLOAAAAPSSP-ATPYSGVGGAIYGERKV 847  
DB 107 AALSVTTDKNLSTGFSS-----LTFLAAPSSVITTPSGKGA----- 144  
QY 848 TFSQSGTCQFSGNQAINNPS-----QSSINVQGGAIYAKTSLISGSSDAGTSYIFSGNS 903  
DB 145 ---KCGGDLTF-----DNGTILFKQDYCEENGGAISTK-NLSLKNSTGSGS--FEGN- 191  
QY 904 VSTGKSQTTGOIAGGAIYPTVTL---NCPATFSNNNTASITPKTSSDGGSGNSIKDTI 960  
DB 192 ---KSSATGKGGKGAICATGTDIINNTAPLFSNNIAEA----- 228  
QY 961 GGAIAGTATLTSVGRFSQNTADLGAAGTANANTPSATSGNSITEKITLENGSFTF 1020  
DB 229 GGAINST-----GNCTITGNTS-----LVFSENSVTATAG-----NG----- 260  
QY 1021 ERNOANKGAIYSPS-VSTKGN-NITENONTSHDGSALYFTKDATIES--LGSVLFTGN 1076  
DB 261 -----GALSGDADVTISGNSQVTFSGNQAVANGGAIY-ARKLTLASGGGGGGSFSNN 311  
QY 1077 NVTATQASSATSGQNTANYGAAIFGDPGTQSSQTDAILTLASSGNITFSNNSLQNN 1136  
DB 312 IVQGT-----TAGNGGAI-----SILAAGECSLSAEAGDITFNGNAIVAT 351  
QY 1137 QGDTPASFCSTAGVYKL-SLOAAKGTISFPDCHVTSTKGTGSGNQVYETLDINKEE--- 1193  
DB 352 TPQTTKRNSIDIGSTAKITNLRAISHSIFFYDPI-TANTAADST-----DTLNLKADAG 406  
QY 1194 NSNPYTGTVFSESE-LHENKSIYIPON-----ALLHNGTLVLKTEKTELHVVSFEQKEG 1244  
DB 407 NSTDYSGSIYVSGEKLSEDEAKVADNLSTLTKOPVTLTAGNLVLKRGVTLDTKGFQTAG 466  
QY 1245 SKLIMEPGAVLSNQNTANGALINGLTIDLDSMGTPQAGEIFSPPELRIVATTSSASGGS 1304  
DB 467 SSVINDAGTTL---KASTEETVLTGLSIPVDSLGECK-----KVVIASASAS--- 510  
QY 1305 GVSSSIPTNPKRISAAPVSGSAATTPTMSKNKVFLLTDLIDPNGNFYQNPMLGSLDV 1364

DB 511 -----KNVALSGPILLLDNQGNAYENHDLGKTQDF 540  
QY 1365 PLIKL-----PTNTSDYQVYDLTSLGDLFPQKGYMGWTFLD-----SNPQTGKLQARWTF 1414  
DB 541 SFVQLSALGTATTTDPAVPTVATPHY---GYQGTWGTWVDVDTASTPKTKATLATLT- 596  
QY 1415 DTYRRWYIP---RDNHFYANSILSGNSMIVVKOGLINMLNNAFDDIAYNNFW/SGV 1471  
DB 597 ---NTGYLPNPEROGPLVPNSLWGSFSDIQAI-QGVIERALTLCSD-----RGFWAAGV 647  
QY 1472 GTFLAQQGTPLSEEFYSYSGTSVAIDAKPRQDFILCAAFSKIVGKTK--AIKKMHVYFH 1529  
DB 648 ANFLDKDKKGRKRYKSGGYAIGGAAQTCSENLSFAFCQLFGSDKDFLVAKNHTDTY 707  
QY 1530 KGSEYSYQASVYGGKFLYFLLNKOHG-WA-LPFLIOGVVSYGHIKHDTTTTLYPSIHENRK 1587  
DB 708 AGAFYQHTTECSG-FIGCLLDKLPGSWSHKPLVLEGQLAYSHVSNDLKTKYTAYPE-VK 765  
QY 1588 GWMEDLGLADLRISMCLKPEKSKRITVYGELEYSSIRKQFTTEIDYDPRHFDICAY 1647  
DB 766 GSWGNNAFNMLGASSHSYPEYLHCFTVAPYIKLNLTVIRODSFSEKGTGGRSFDJNSL 825  
QY 1648 RNLSLPVGCAGEAIMNCNMLKALAYMPSIYRNNPVCKYRVLSSNEAGQVIGVPT 1707  
DB 826 FNLSLPVIGVKEE-KFSDCNDFSYD-LTSLSYVDLIRNDPKCTTALVISGASWETAYAINLA 883  
QY 1708 RTSARAESTQLYLGFPFTLYGNTIDV 1735  
DB 884 RQALQVRAGSHYAFSPWFVLGQFVEV 911

Search completed: October 25, 2001, 08:42:40  
Job time: 214 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 08:36:01 ; Search time 28.96 Seconds  
(without alignments)  
4608.350 Million cell updates/sec

Title: US-09-454-684A-180  
Perfect score: 8942  
Sequence: 1 MKWLSATAVFAAALPSVSGF.....IDVGMVTLQMTSCGARMIF 1752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8827.5	98.7	1751	2 G71518	hypothetical prote
2	6488.5	72.6	1672	2 C81675	polymorphic membra
3	3550	39.7	1770	2 A71517	hypothetical prote
4	3205	35.8	1460	2 D81675	polymorphic membra
5	3112.5	34.8	1723	2 E72067	polymorphic membra
6	3111.5	34.8	1732	2 H86557	polymorphic membra
7	3111.5	34.8	1732	2 C81601	polymorphic membra
8	932.5	10.4	1407	2 B72078	polymorphic outer
9	823.5	9.2	1276	2 B86546	polymorphic outer
10	823.5	9.2	1276	2 C81591	polymorphic membra
11	743.5	8.3	1609	2 A86611	probable outer mem
12	743.5	8.3	1609	2 D72013	polymorphic membra
13	674	7.5	1531	2 H71468	probable outer mem
14	665.5	7.4	930	2 D86546	polymorphic outer
15	665.5	7.4	930	2 A81591	polymorphic membra
16	661.5	7.4	930	2 D72078	polymorphic outer
17	655.5	7.3	928	2 G86546	polymorphic outer
18	655.5	7.3	928	2 B81591	polymorphic membra
19	622.5	7.0	1520	2 A81731	polymorphic membra
20	577	6.5	2232	2 T34434	hypothetical prote
21	555.5	6.2	936	2 C72078	polymorphic outer
22	553.5	6.2	936	2 C86546	polymorphic outer
23	553.5	6.2	936	2 B81591	polymorphic membra
24	530.5	5.9	922	2 E86491	polymorphic outer
25	530.5	5.9	922	2 F81539	polymorphic outer
26	528.5	5.9	922	2 F81539	polymorphic membra
27	517.5	5.8	949	2 F81591	polymorphic outer
28	516.5	5.8	928	2 H86546	polymorphic outer
29	516.5	5.8	928	2 D72077	polymorphic outer

30 516 5.8 975 2 F71518 hypothetical prote  
31 509 5.7 772 2 H86492 pmp\_3 [imported] -  
32 506 5.7 841 2 E72130 polymorphic membra  
33 504 5.6 1013 2 G71460 probable outer mem  
34 495 5.5 928 2 E86546 polymorphic outer  
35 495 5.5 928 2 B72077 polymorphic membra  
36 478 5.3 986 2 B81675 polymorphic membra  
37 470 5.3 973 2 B86547 polymorphic outer  
38 470 5.3 973 2 F72076 polymorphic outer  
39 470 5.3 995 2 C81593 polymorphic membra  
40 466 5.2 1016 2 H71460 probable outer mem  
41 461.5 5.2 978 2 B81593 polymorphic membra  
42 461.5 5.2 978 2 C86547 polymorphic outer  
43 456.5 5.1 978 2 G72076 polymorphic outer  
44 452 5.1 2249 2 A41477 190K surface antig  
45 448.5 5.0 987 2 H81722 polymorphic membra

#### ALIGNMENTS

RESULT 1

G71518

hypothetical protein pmpB - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: G71518

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t

A:Reference number: A71570; MUID:99000809

A:Accession: G71518

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1751 <ARN>

A:Cross-references: GB:AE001314; GB:AE001273; NID:g3328833; PIDN:AC68010.1; PID:g332

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: pmpB

Query Match 98.7%; Score 8827.5; DB 2; Length 1751;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1732; Conservative 8; Mismatches 10; Indels 3; Gaps 2;

Qy 1 MKWLSATAVFAAALPSVSGFCPEPEKLNFSRVETSSSTFTTETICEAGAETVSGNASF 60

Db 1 MKWLSATAVFAAALPSVSGFCPEPEKLNFSRVETSSSTFTTETICEAGAETVSGNASF 60

Qy 61 TKFTNPTTDTTPTNSNSSSSGSETASVSDSDSTTTTTPDKGGGAFYNAHSGVLSFMT 120

Db 61 TKFTNPTTDTTPTNSNSSSSGSETASVSDSDSTTTTTPDKGGGAFYNAHSGVLSFMT 120

Qy 121 RSGTEGSLSEIKMTGEGGAIIFSQGLLELTDLTSLTQNNLSQLSGGAIFGSGTSLSG 180

Db 121 RSGTEGSLSEIKMTGEGGAIIFSQGLLELTDLTSLTQNNLSQLSGGAIFGSGTSLSG 180

Qy 181 ITKATFSCNSAEVPAVPKPTPEKAOTASSETSGSSSSGNDVSSPSSRAEPAANLQS 240

Db 181 ITKATFSCNSAEVPAVPKPTPEKAOTASSETSGSSSSGNDVSSPSSRAEPAANLQS 240

Qy 241 HFICATATPAAQTDSTETPSHKPGGGGAIYAKGDLTIADSQEVLFISINKATDKGAIFA 300

Db 241 HFICATATPAAQTDSTETPSHKPGGGGAIYAKGDLTIADSQEVLFISINKATDKGAIFA 300

Qy 301 EKDVSPENITSLKVQTNGAEEKGAIYAKGDLISQSSKOSLFNSNYSKOGGALYVE3GI 360

Db 301 EKDVSPENITSLKVQTNGAEEKGAIYAKGDLISQSSKOSLFNSNYSKOGGALYVE3GI 360

Qy 361 NFQDLEIRIKYNKAGTFFETKKTITLPSLKAQASAGNADAWASSPSSQSGCATTVSDS 420

Db 361 NFQDLEIRIKYNKAGTFFETKKTITLPSLKAQASAGNADAWASSPSSQSGCATTVSDS 420

Qy 418

Db 418

QY 421 SSGSDTSERVPVAKGGGLYTDKNLSITNITGTHIEIANANKATDVGGAIVKGTGLTCEN 480  
|||||  
Db 419 SSGSDTSERVPVAKGGGLYTDKNLSITNITGTHIEIANANKATDVGGAIVKGTGLTCEN 478  
|||||  
QY 481 SHRLOFLKNSDKQGGGIYGEDNITLSNLTGKTLFQENTAKEEGGLFIKGTDKALMTWG 540  
|||||  
Db 479 SHRLOFLKNSDKQGGGIYGEDNITLSNLTGKTLFQENTAKEEGGLFIKGTDKALMTWG 538  
|||||  
QY 541 LDSFCLINNTSEKHHGGGAFVTKELISOTVTSDETTPGTPVHGETVITGNKSTGGGGV 600  
|||||  
Db 539 LDSFCLINNTSEKHHGGGAFVTKELISOTVTSDETTPGTPVHGETVITGNKSTGGGGV 598  
|||||  
QY 601 CTKRLALSNLQISISISGNSAENGGAHTCPDSPPTADTAEPAAASAATSPKSAP--VS 659  
|||||  
Db 599 CTKRLALSNLQISISISGNSAENGGAHTCPDSPPTADTAEPAAASAATSPESAPVVS 658  
|||||  
QY 660 TALSTPSSSTVSSLLLAASQAASPATSNKETQDPNADTDLIDYVVDTTISKNTAKGG 719  
|||||  
Db 659 TALSTPSSSTVSSLLLAASQAASPATSNKETQDPNADTDLIDYVVDTTISKNTAKGG 718  
|||||  
QY 720 GIYAKKAKMSRIDOLINISENSATEIGGGICCKESLELDALVSLVTENLVGEGGGLHAK 779  
|||||  
Db 719 GIYAKKAKMSRIDOLINISENSATEIGGGICCKESLELDALVSLVTENLVGEGGGLHAK 778  
|||||  
QY 780 TVNISNLKSGFSSFNKANSSTGVATTASAPAAAAAASLQAAAAAPSPATPTVSYGVG 839  
|||||  
Db 779 TVNISNLKSGFSSFNKANSSTGVATTASAPAAAAAASLQAAAAAPSPATPTVSYGVG 838  
|||||  
QY 840 GAIYKEKVTFCOCSTCQFSGNAIDNNPSSSLNVQGGAIYAKTSLSIGSDAGTSYIF 899  
|||||  
Db 839 GAIYKEKVTFCOCSTCQFSGNAIDNNPSSSLNVQGGAIYAKTSLSIGSDAGTSYIF 898  
|||||  
QY 900 SGNVSSTCKSQTGTGOIAGGAIYSPVTIVLNCPCATFSNNATPATKTSSEDESSGNSIKDT 959  
|||||  
Db 899 SGNVSSTCKSQTGTGOIAGGAIYSPVTIVLNCPCATFSNNATPATKTSSEDESSGNSIKDT 958  
|||||  
QY 960 IGGAIAGTAITLSGVSREFSGNADLGAAGTLANANTPSATSGSONSITEKITLENGSEFI 1019  
|||||  
Db 959 IGGAIAGTAITLSGVSREFSGNADLGAAGTLANANTPSATSGSONSITEKITLENGSEFI 1018  
|||||  
QY 1020 FERNOANKRGAIYSPSVSIKGNITFNQNTSTHDSAIYFTKDIATIESLGSVLTFTGNVY 1079  
|||||  
Db 1019 FERNOANKRGAIYSPSVSIKGNITFNQNTSTHDSAIYFTKDIATIESLGSVLTFTGNVY 1078  
|||||  
QY 1080 ATQASSATSGQNTNTANYGAIIFGDPGTTQSSQTDAILTLASSGNITFSNNSLQNNQGD 1139  
|||||  
Db 1079 ATQASSATSGQNTNTANYGAIIFGDPGTTQSSQTDAILTLASSGNITFSNNSLQNNQGD 1138  
|||||  
QY 1140 TPASKFCSIAGYVVLKSLQAAGKKTISFPDCVHTSTPKTGTSTQNVYETLDINKEENSNPY 1199  
|||||  
Db 1139 TPASKFCSIAGYVVLKSLQAAGKKTISFPDCVHTSTPKTGTSTQNVYETLDINKEENSNPY 1198  
|||||  
QY 1200 GTIVFSSSELHENKSIYIPONAILHNGTIVLKEKTELHVVSPEQEGSKLIMPEGAVLSNQ 1259  
|||||  
Db 1199 GTIVFSSSELHENKSIYIPONAILHNGTIVLKEKTELHVVSPEQEGSKLIMPEGAVLSNQ 1258  
|||||  
QY 1260 LANGALAINGLTIDLSSMGTPOAGEIFSPPELRIVATTSSASGGGVSSSIPTNPKRTISA 1319  
|||||  
Db 1259 LANGALAINGLTIDLSSMGTPOAGEIFSPPELRIVATTSSASGGGVSSSIPTNPKRTISA 1318  
|||||  
QY 1320 AVPSGSAATPTMSSENKVELTDLIDPNGNFYONPMLGSDLDVPLIKLPTNTSDVOVY 1379  
|||||  
Db 1319 AVPSGSAATPTMSSENKVELTDLIDPNGNFYONPMLGSDLDVPLIKLPTNTSDVOVY 1378  
|||||  
QY 1380 DLTLSGDLFPQGYMGVWTLDNSPOTGKLOARWTFDTPYRWWYIPRONHFFYANSLILGSON 1439  
|||||  
Db 1379 DLTLSGDLFPQGYMGVWTLDNSPOTGKLOARWTFDTPYRWWYIPRONHFFYANSLILGSON 1438  
|||||  
QY 1440 SMIVVYKQGLINMLNNAFFDDIAYNNFVWSVGVTFLAQQGTPLSEEFYSYSGTSVAIDA 1499  
|||||  
Db 1439 SMIVVYKQGLINMLNNAFFDDIAYNNFVWSVGVTFLAQQGTPLSEEFYSYSGTSVAIDA 1498  
|||||  
QY 1500 KPRODFILGAAFSKIVGKTKAIKKMHNTYHKGSEYSYQASVYGGKFLYLLNKOHGWALP 1559  
|||||

Db 1499 KPRODFILGAAFSKIVGKTKAIKKMHNTYHKGSEYSYQASVYGGKFLYLLNKOHGWALP 1558  
|||||  
QY 1560 FLIOGVWSYGHKHDHTTTLPSIHERNKGWEDLGLWADLRISMDLKEPSKDSKRITVY 1619  
|||||  
Db 1559 FLIOGVWSYGHKHDHTTTLPSIHERNKGWEDLGLWADLRISMDLKEPSKDSKRITVY 1618  
|||||  
QY 1620 GELEYSSIRQOFTEIDYDPRHFDCCAYRNLSLPVGCACAVEGAIMNCILMYNKLALAYMP 1679  
|||||  
Db 1619 GELEYSSIRQOFTEIDYDPRHFDCCAYRNLSLPVGCACAVEGAIMNCILMYNKLALAYMP 1678  
|||||  
QY 1680 SIYRNNPVCKYRVLSSNEAGQVIGCVPTRTSARAEYSQLYLGPFWTLYGNVYIDVGMYT 1739  
|||||  
Db 1679 SIYRNNPVCKYRVLSSNEAGQVIGCVPTRTSARAEYSQLYLGPFWTLYGNVYIDVGMYT 1738  
|||||  
QY 1740 LSQMTSCGARMIF 1752  
|||||  
Db 1739 LSQMTSCGARMIF 1751  
|||||  
RESULT 2  
C81675  
polymorphic membrane protein B/C family TC0694 [imported] - Chlamydia muridarum (stra  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: C81675  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
A: C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: C81675  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1672 <FET>  
A:Cross-references: GB:AE002338; GB:AE002160; NID:g7190724; PIDN:AAF39510.1; PID:g719  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0694  
  
Query Match 72.6%; Score 6488.5; DB 2; Length 1672;  
Best Local Similarity 73.0%; Pred. No. 1.9e-301;  
Matches 1292; Conservative 145; Mismatches 214; Indels 119; Gaps 20;  
  
QY 1 MKWLSATAVPAALVPSVSGFCFPEPKELNFSRVETSSSTTTTETIGEAGAEYIVSGNASF 60  
|||||  
Db 4 MKWLSATAVPAALVPSVSGFCFPEPKELNFSRVETSSSTTTTETIGEAGAEYIVSGNASF 63  
|||||  
QY 61 TKFTNI----PTTD---TTTPTNSNSSSSSGE---TASVSESDSDSTTTTDPKGGGAFY 109  
|||||  
Db 64 TNFTNIPVKKPTTDDSSSTSTPTSSAVDPTEKIVPASSSSSPNSGDTSATDPKGGGAFY 123  
|||||  
QY 110 NAHSGVLSFMTRSCTEGSLTISEIKMTGEGGAIFSQGELLFTDLTSLTQNNLSQSGA 169  
|||||  
Db 124 NEHSGILSFMAARSGVEGSLTSLNKMKTGGGAIYSQGELLFTDLTGLTIQNNLSQSGG 183  
|||||  
QY 170 IFGGSTISLGGITATFSCNSAEVPAVPVKKTEPKAQTASETSGSSSSSGNDSVSSPSS 229  
|||||  
Db 184 IFGGSTISFSGINQATFSNTAEV---VPEETTPNPNGTQTTSQ-----PSPTS- 231  
|||||  
QY 230 RAEPAAANLQSHFCATATPAATQDTSTPESHKPGSGAIYAKGDLTIADSQEVLFSIN 289  
|||||  
Db 232 -----KVQSLFYSSSTQANGCADSQTPSHKPGSGAIYATGDLTISDSQEIFSVN 284  
|||||  
QY 290 KATKDGGAIFAEDVSNFENITSLKVQTNABEKGGAIVAKGDLSTQSSKQSLFNSYSKQ 349  
|||||  
Db 285 KASDKGGAIFAEDVSNFENITSLKVQTNABEKGGAIVAKGDLSTQSSKQSLFNSYSKQ 344  
|||||  
QY 350 GGGALYVEGGINFQDLEERIKYNKAGTFETKTLPSLKAQASAGNADANASSPQSGS 409  
|||||  
Db 345 GGGALYIEGNVDFKDLLEERIKYNKSGTFETKTKVTLSPLEAQTNKSSVTA----- 394  
|||||

QY 410 GATTVSDSGSSGSDSTSETVPVTA KGGGLYTDKNLSITNITGIEIANNKATDVGGG 469  
Db 395 -----ASQSGPNVTPTTPVTA KGGGLYTERKLSNITGIEIETNNKATDVGGG 445  
QY 470 AYVKGTLTCNSHRLQFLANKSSDKGGGIYGEDNITLSNLTGKTLFQENTAKKEGGGLFI 529  
Db 446 AYVKGTLTCNSHRLQFLANKSSDKGGGLYEDITLSNLTGKTLFQENTAKKEGGGLYI 505  
QY 530 KGTDKALTMTGLDSFCLINNTSEKHGGGAFVTKKEISQYTSQVETIPGTPVHGETVITG 589  
Db 506 QGDDKLTMTGLDSFCLIDNTSATHGGGAVTKKEISQYTSQVDEEFPGITPVHGETIISG 565  
QY 590 NKSTGGGGVCTKRIALSNLQISISGNSAENGGAHTCPDPSFTADTAEOPAASAA 649  
Db 566 NKATGGGGGVCTKHLVLSNLTQISISFASENGGAHTCPDNFP----- 611  
QY 650 TSTPKAPVSTALSTPSSSTVSSVSLTLAASSQASPATSNKETODPNADTDLLDYVVDTT 709  
Db 612 -----APTA---STFSIN-----QTAAPKDDKDFLDYVVSST 641  
QY 710 ISKNTA-KKGGGIYAKKAKMSRIDOLNISSENSATEIGGGICCKESLELDALVLSVTENL 768  
Db 642 IDNKATKKGAGVYAKKALSRIDELNISDNAAQETGGGFCCTESLELDTIASLSVTKNL 701  
QY 769 VKEGGGLHAKTVNLSKGFSEFSNNKANSSTGVATTASAPAAAASLQAAAAAPSS 828  
Db 702 AKREGGGLHAKTLNLSKLSGLSFSNNTANSSTGVATTATTSQSPTVSSFLPRATAGSS 761  
QY 829 PA----TPYSGVVGAIYGEKVFESQCGTCQFSGNOAIDNNPSSOSSLNVQGGAIYAKT 884  
Db 762 PAPAQTTPYAGVVGAIYGETVFSKCSGLCQFTENSAIDNTPSPSLNVQGGAIYAKT 821  
QY 885 SLSIGSDAGTYSIFSGNSYSTGSKSTQGIAGGAIYSPVTILNCPATFSNNTASITATPK 944  
Db 822 SLSIEADPSTSVFSGNSYSTGKAQTTGQIAGGAIYSPSVTLNCQTVFSGNSASMAT-- 879  
QY 945 TSSEDSGSSNSIKDITGGIAGTAITLGSVRSFGNNTADLGAAGTLANANTPSATSGSQ 1004  
Db 880 ---TNPSPGTSPPKDTGGIAGTITISLSKTSHPSENTADLGAAGITL-----SGGSS 928  
QY 1005 NSTEKITLNGSFIFERNOANKRGAIYSPSVSIKGNNTFNONTSHDGSALYFTKDAT 1064  
Db 929 SNTSEKITLNGSFTEKKANKARGIYAPSVSIKGNNTFNONTSHDGSALYFTKDAT 988  
QY 1065 IESLGSVLEFTGNVNTATQASS-ATSQONTANYGAAIFGDPCTQTSSQTDAILTLIASS 1123  
Db 989 IESLGSVLEFTGNVNTAEQASSTATGQTTNTYGAIFGDPCT---QTDTTLKLIASS 1045  
QY 1124 GNTTFNSNLSQNNQDTPASKFCISAGYKLSLQAAGKGTISFFDCVHTSTKKTGTQNV 1183  
Db 1046 GNTTFNSNS-QNTATNPATPKFCISGYVKLTLLQAAGKGTISFFDSIRTSTKKTGQAQNS 1104  
QY 1184 YETFLDNKENSNPYGTIVFSELHENKSYIPQNAIHLNGTTLVLEKTEHLVVSFEQKE 1243  
Db 1105 YETFLDNKNTENSNTYAGTVLFSELHELVKSYVQNVVHLNGTTLVLEKTEHLVVSFEQKE 1164  
QY 1244 GSKLIMEPGAVLSNQNIANGALINGLTIDLSSMGTPQAGEIFSPPELRIVATTSASGG 1303  
Db 1165 GSKLIMEPGAVLSNQNIANGALINGLTIDLSSLGAPQAGEIFSPPELRIVATTSNSGGG 1224  
QY 1304 SGVSSSIPNPKRISAAVPSGSAATPTPMSKNVFLTGDLTLDPNGNFYQNPLGSDL- 1362  
Db 1225 GGVGVV-TASKNLSAASPT-VAATNPTMADNKVFTLTGALTLDLPNGNFYQNPLGSDLT 1282  
QY 1363 DVPLIKLPTNTSDVOYDITLSGDLFPQKGYMGCTWTPLDSNPQTGKLQARWTRDTPYRWVY 1422  
Db 1283 DVPLIKLPTTANQVDVNSLTLSGDLSPKGYGTGTWTLPNDPQTKVYANWKFDMYRWY 1342  
QY 1423 IPRDNHFIANSILGSONSMIVVKQGLINNLNNAFDDIAYNNFWVSGVGTFLAQQGTPL 1482  
Db 1343 IPRDNHFIANSILGSONSMIVVKQGLINNLNNAFDDIAYNNFWVSGVGTFLAQQGTPL 1402  
QY 1483 SEEFYSYSRGTSVAIDAKPRQDFILGAAFSVKGTAKIKKMHNYFHKGSEYSYQASVYG 1542

Db 1403 SEEFYSYSRGTSVAIDAKPRDFILGAAFSKMVGRTKAIKKVHNSHKGSEYSYQASVYG 1462  
QY 1543 GKFLYFLLNKHQGWALPFLQGVVSYGHIKHDITTLVPSIHERNKGDMEDLGLWDLARIS 1602  
Db 1463 GKFLYFLLNKHQGWALPFLQGVVSYGHIKHDITTLVPSIHERNKGDMEDLGLWDLARIS 1522  
QY 1603 MDLKEPSKDSKRTIYVGELEYEYSIRQKQFTEIDYDPRHFDCCAYRNLSLPVGCAVBGAI 1662  
Db 1523 MDVKEPSKRSSKRVALYGELEYEYSIRQKSFTEIDYDPRHFDCCAYRNLSPMGCYFBGAI 1582  
QY 1663 MNCNLMYKALAYPISYIYRNPNVCKYRVLSSNEAGQVTCGVPTRTSARAESTQYILG 1722  
Db 1583 MSYDILMYKLSLAYPISYIYRNPNVCKYRWYVLSNETSKVYGVPTRTSARAESTQYILG 1642  
QY 1723 PFWTLGNYTIDVGMVTLTSSOMTSCGARMIF 1752  
Db 1643 PFWTLGNYTIDVGMVTLTSSOMTSCGARMIF 1672  
RESULT 3  
A71517  
hypoetical protein pmpC - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: A71517  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:990000809  
A:Accession: A71517  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1770 <ARN>  
A:Cross-references: GB:AE001315; GB:AE001273; NID:g3328842; PIDN:ALC68011.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: pmpC  
Query Match 39.7%; Score 3550; DB 2; Length 1770;  
Best Local Similarity 39.9%; Pred. No. 2.6e-161;  
Matches 812; Conservative 259; Mismatches 416; Indels 548; Gaps 39;  
QY 1 MKWLSATAVFAAALPSVSGRCFPPEK---ELNFSRVETSSSTFTTETICEAGAEYIVSG 56  
Db 1 MKFMSATAVFAAALSSVTEASSIQDQIKNTDCNVKLGISTSAFTDMLADNTEYFAAD 60  
QY 57 NASFTKFTNPTTDTTPTTNSNSSSSGGETASVSESDSTTTTPDPKGGGAFYNAHSGVL 116  
Db 61 SVSFYDF---STSSRLPRKHLSS----- 81  
QY 117 SFWTRSGTEGSLTSLSEKMTGEGGAIFSQGELLFTDLTSLTIQNLSQLSGAIFGGSTI 176  
Db 82 -----SEASPTTEG----- 90  
QY 177 SLSGITKATFCNSAEVPAVPKPTPEPKAQTASSETSGSSSSGNDSSVSPSSSRAEFAA 236  
Db 91 -----VSSSSSG----- 97  
QY 237 NLOSHFICATATPAQDDETSTPSHKPGSGGAIYAKGDLTIADSQ----- 282  
Db 98 -----ETDEKT---EEELNGGIIYAREKLTISESDSLSNQSTELHDNS 139  
QY 283 -----EVLFSINKATKDGCAIFAEXDVSFENITSLKVQTNGAEEKGAIYAKGDLISIQS 336  
Db 140 IFEGEIVFDHRVAKNGGAIYGEKEWFENIKSLLEVNIYAVEKGGSYAKERSVLEN 199  
QY 337 SKOSLNSNSKOGGALYVEGGINGFODLEEIRIKYNKAGTFTTKT----- 383  
Db 200 VTEATFSSNGCEGGGIYSEQDMLISDCNNVHFGQNAAGATAVKQCLDEEMIVLIAECV 259  
QY 384 -----TLPSLKAQASAGNADAWASSSPQSGGATTVSDSGSDSTSETVPY 434

Db 260 DLSLSDTLDTSTPTEQTESNGNODG-----SSEFEDTQVSESPSTSPD-----V 306  
Qy 435 TAKGGGLTYDKNLSTNTLTGIIETIANNKATDVGGAYVKGTLTCENSHRLQFLKNSSDKQ 494  
Db 307 LKGGGIIYTERSLTITGTLIDFVSNATDSGAGVFTKENLSCNTNSLQFLKNAG-- 364  
Qy 495 GGGIYGEDNITLSNLTGKTLFOENTAKEEGGGLFKGTDKALMTGTLGDSFCLINNTSEKH 554  
Db 365 -----QH 366  
Qy 555 GGGAPVTKESOT--YTSVETIPGTPVHGTVITGNKSTGNGGCVCTKRLALSNLQSI 613  
Db 367 GGGAVTQMTSVTNTSITP---PLIGEVIFSEN-TAKGHGGGICTNKLSLNLTAV 422  
Qy 614 SISGNSAENGGAHTCPDSPPTADTAEQ--PAASAATSTPK---SAPVSTALSTPSS 668  
Db 423 TLTNKAKESGGAIFDTLJASIPITDPESSTPSSSPA-STPEVVASAKINRFFASTAKP 481  
Qy 669 TVSSLTLLAASQAASPATSNKQTPDNADTOLLIDYVDVTTISKNT-AKKGGGIYAKKAK 727  
Db 482 AAPSLT-----EASDQTDOTSTDSNIDIVSINILNVALNQNTSAKGGAIYKRAK 536  
Qy 728 MGRIDOLNISENATEIGGICKESLELDALVLSVSTENLVKKEGGGLHAKTVNISNLK 787  
Db 537 LSRINNELSGNSQDVGGLCTESVEFDAIGSLLSHYNSAAKEGGAIHSTKVTLSNLK 596  
Qy 788 SCFSF-----SNKANSST----- 802  
Db 597 STFTFADNTVKAIVSTPEAPEIPVPEGEESTATEDPNSNTEGSSANTNLEGSQGTAD 656  
Qy 803 ----- 802  
Db 657 TGTGVDNNEQDTSDTGNAESSEQLQDSTQSENEENTLPSNIDQSNENTDESSDHTBEI 716  
Qy 803 -----GVATTASAP-----AAAAAALQAAAAAPSSP----- 829  
Db 717 TDESVSSESGSTPQDGGAAAGAPSGDQSIANACLAKSYAAASTDSSPVNSSGSEE 776  
Qy 830 -----ATPTYGVGVGAIYGEKVIFS 850  
Db 777 PVTSSSDSDVTASSNPDPSSSGDSAGDSEPEPEAGSTTETLLIGGGAIYGETVIE 836  
Qy 851 QCSGRQCFSGNOAIDNPP--SOSSLNVGGAIYAKTSLISGSDAGTYIFSGNSVSTGK 908  
Db 837 NPSGGIFSGNKAIDNTTEGSSSKSDVLGGAVYAKTLNLDGSSRRVTTFSGNVSS-- 894  
Qy 909 SOTTGQIAGGAIYSPVTNLNCPATFSNNTASITATPKTSESDGSSGNSKDTIGGAIAGT- 967  
Db 895 QSTTGQVAGGAIYSPVTIATPVVFSKNSA-----TNNANNTTDTQRKDTFGGAIGATS 948  
Qy 968 ATTLSGVRSFGSNTADLGAAGICTLANANTPSTATSGSONSITEKITLNGCSFIFERNQANK 1027  
Db 949 AVLSGGGAHLENADLGAAGLV-----PGTON--TETVKLESGSYFFKRNKALK 997  
Qy 1028 RGAIYSPSVSIKGNNTNONTSTHDGSAIYFTKDATIESLGSVLFTGNVNTATQASSAT 1087  
Db 998 RATIYAPVSIKAYTATNQNSLEEGSAIYFTKEASIESLGSVLFTGNVLT-LSITT 1056  
Qy 1088 SQONTNTA-----NYGAAIFGDCPTQSSQTDAL-UTLASSGNITFSSNLSQNNQCDTPA 1142  
Db 1057 EGTPTATSGDVTKYGAAIFGQIASSGQTDNLPLKLASGGNICFRNNEYRPTSSDTGT 1116  
Qy 1143 SKFCSTAGVVKLSLQAAKGKTIISFFDCVHTSPKKTGSTQNVYETLIDINKEENS---NPY 1198  
Db 1117 STFCSTAGDVKLTMOAAGKTIISFFDAIKSTPKTKGTQATAYDTLIDINKSEDSFTVNSAF 1176  
Qy 1199 TGTIYFSSSELHENKYIPONAILHNGTLVLKKEKTELHVYVFPQKEGSKLIMEPGAVLSNQ 1258  
Db 1177 TGTILFSSSELHENKYIPONVVLHSGSLVKLPNTLHVISFPQKEGSSILVWTPGVSLSNQ 1236  
Qy 1259 NTANGALAINGLTIDLSSM---GTPQAGBIFSPPELRIVATTTSSAGG----- 1303  
Db 1303 ----- 1303

Db 1237 TVADGALVINNMITIDLSSVEKNIAE-GNIFTPELRIIDTTTGGSGGTPSTDSSESNQS 1295  
Qy 1304 -----SGVSSSIP-----TNPKRISAAVPSGSAATTP---TMSENKVF 1338  
Db 1296 DDTBEONNDASNOGESANGSSSPAAVAHAHTSRFRNFAAATAATPTTPTATTATTSNOVI 1355  
Qy 1339 LTGDLTLLIDPNGNFYQNPMLGSDLDVPLIKLPTNTSDQVYDVLTLGSLDFPKQGYMGWT 1398  
Db 1356 LGGEIKLIDPNTGTFQNPALRSDDQISLLVLPDSSKMQAQKIVLTGDIAPKQGYTGLT 1415  
Qy 1399 LDSNP-QTGKLOARWTFDTYRRWYIIPRDNHFYANILGSONSMIVVKQGLNNMLNAR 1457  
Db 1416 LDPDQLQNTGISVLKFDYSRQWYAVPRDNHFYANILGSONMLVTVRQGLNDKNLAR 1475  
Qy 1458 FDDIAYNNFWGSGVGTFLAQQGTPLSEEFYSYSGTSAIDAKPRQDFILGAASKIVGK 1517  
Db 1476 FEESYNNLWISGLTMLSUGVPTPTESEFTYISRGASVALDAKPAHDVIVGAASFMTGK 1535  
Qy 1518 TKAIKMHNYPHKGSEYSYQASVYGGKFLYFLNKKHGWALPFLTQGVYVSGYGHKHDITT 1577  
Db 1536 TKSLLKRENNYTHKGSEYSYQASVYGGKPFHFVINKKTEKSLPLLQGVISYGIKHDVT 1595  
Qy 1578 LYSIHERNKGDWEDLGLWADLRISMDLKEPSKSSKRITVYGELEYSSIRKQFTEIDY 1637  
Db 1596 HYPITRERNKGEWEDLGLWALTALRVSSVLRTPAQGDKRITVYGELEYSSIRKQFTEY 1655  
Qy 1638 DPRHFDCCAYRNLPLVPCAVEGAIMNCNILMYNKLAYMPSIYRNNPVCYKRVLSNE 1697  
Db 1656 DRYFDNCTYRNLAIIPMGLAFEGELSGNDILMYNRFVAYMLSIYRNSPTCKYQVLSGE 1715  
Qy 1698 AQVICGVPYRTSARSTYQLYGLFPFWTLGNYTIDYGMVTLTSLQMTSCGARMIF 1752  
Db 1716 GGEIICGVPTNRNARGEYSTQLYGLPLWTLGYSYTIADAHTLAHMMNCGARMTF 1770

RESULT 4  
D81675  
polymorphic membrane protein B/C family TC0695 [imported] - Chlamydia muridarum (stra  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: D81675  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: D81675  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1460 <TET>  
A:Cross-references: GB:AE002338; GB:AE002160; NID:g7190724; PID:AAF39511.1; PID:g719  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0695

Query Match 35.8%; Score 3205; DB 2; Length 1460;  
Best Local Similarity 39.3%; Pred No. 5.5e-145;  
Matches 741; Conservative 218; Mismatches 366; Indels 562; Gaps 33;

Qy 1 MKWLSATAVFAALPSVSGVCFPP-----EPKELNFRSVETSSSTTTTETIGEAGAIYVSG 56  
Db 1 MKFLSATAVFAAALPSITSASSVESQIETKDLNSRSTGSSSSQSSTTEIIPENGAEVYVSG 60

Qy 57 NASFTKFTNIP-----TSTDTPPNS 77  
Db 61 DVSFSDFNIPPEAEETLAISHKEQNPNEVLSSEENHQASQDSQNTENASEGNSPSE 120

Qy 78 NSS-SSSGETAS-----VSESDSDSTTTP----- 100  
Db 121 NTNQSSSTTETESITTDQEQNDNESAAVPTTETATAMRLPSYHLQTESLVEGATEEDO 180

Qy 101 -----DPKGGGAFYNHSGVLSFMTSTRSGTBGSLTSLSEIKMTGEGGAIFSGELLFTDL 153



Db 61 IVGDITFTFTNIPVPVPTDANDSSNSSKGGSSSGATSLSIRSNLHSDFDFTKDSYL 120  
Qy 87 -----ASVEDSDSTTTTDPKGGGAFY-NAHSG 114  
Db 121 DLYHLFFPSASNTLNPALLSSSSSGSSSSGSASAVVAADPKGGAIFYNEANG 180  
Qy 115 VLSFMTSRGEGSLTLSEIKMTGEGGAFSOGELLFTDLTSLTIONNLSQLSGGAIF--- 171  
Db 181 TLTFTTDSNPGSLTLQNLKMTGDGAAYISKPLVFTGLKLNLTFTGNESQKSGGAAYTEG 240  
Qy 172 -----GG-----STIS 177  
Db 241 ALTTQAIQVAVFTGNTSAGQGAIVYKATLFLNALDSLKFENKTSQAGGGIYTESTIL 300  
Qy 178 LSGITKA-TFSCNSAEVPAVPKPEPKPAQTASETSGSSSSGNDVSSPSSRAEPAAA 236  
Db 301 ISNITSIEFISNKASVPAPPEPTSP-----APSLINSITIDTSLTQTRAASATPAVA 355  
Qy 237 NLOSHFICATATPAQOTDTETSPSHKPGSGGAIYAKGDLTI----- 278  
Db 356 PV-----AAVPTPISTQETA-----GNGGAIYAKOGISISTFKDLTFKSNASVDAT 403  
Qy 279 -----ADSOEV-----LFSINKATDGGAIAPKDVSPENTISLUK 314  
Db 404 LTVDSSTIGESGAIFAADSIOIQOCTGTTLFSGNTANKSGGGIYAVGQVTLIEDIANLKM 463  
Qy 315 QTNAAEKGGAAYAKGDLISQSK-QSLFNSNYSKOGGALYVEGGINFDLEEIRIKYN 373  
Db 464 TNNCKEGGAIYTKALTINGGAILTTFSGNTSTONGGAIYAVGQVTLSDLVEVRFKSN 523  
Qy 374 KAGTETKRTILPSLKAQASAGNADAWSSPQSGSGATTVSDSGSDSSGSDSTSETP 433  
Db 524 KUGNY-----SAPITKA-----ASNTAPVYSSITTAASPAV-----PAAAAA 561  
Qy 434 VT--AKGGGLYTDKNLSITNITGIETIANKATDVGGGAYVKGTLTCENSHRLOFLKNS 491  
Db 562 VTNAAKGALYSTEGTLVTSKYSILSFENECNOGGGAYVTKTQCSDSHRLOFTSNKA 621  
Qy 492 DKOGGIIYGEDNTLSNLTKTLFQENTAKREGGLFIKGTDKALMTGLDSCFLNNITS 551  
Db 622 ADEGGGLYCCDDVTLNLTKTLFQENSSKEHGGGLSL-ASGSLTMTSLESFCLNANTA 680  
Qy 552 EKHGGAFTKREISOYTT---SDVETIPGTVPHGTEVITGNKSTGGNGGVCYTKRLALS 608  
Db 681 KENGGGANVPENILVTFVTPNPNPAPVQPVYGEALVTGNAT-KSGGGIYVTKNAAPS 739  
Qy 609 NLOSTISONSRAENGGAHTCPDSEPTADTAEQAPAAASAASTPKSAPVSTALSTPSS 668  
Db 740 NLSSVTFDQNTSSENG-----ALLTQKAD----- 765  
Qy 669 TVSSLTLAASQAASPATSNKETQDPNADPTLLIDYVVDTTISKNTAK-KGGGIYAKKAK 727  
Db 766 -----KTDSCFYITNVNNTNNTATGNGGIAGGRAH 797  
Qy 728 MSRIDQNLISENATEIGGGICCKESLELDALVLSVTENLVKGGGLHAKTVNISNLK 787  
Db 798 FDRIDNLTVQSNQAKK-GGGVLEDALEKVTIGSVSQTATESGGIYAKDIQLOALP 856  
Qy 788 SGFSFSNNKANSSTGCVATTASAPAAAAASLQAAAAAPSSPATPYISVGVGAIYGE-K 846  
Db 857 GSFTTIDNKVETSLT-----TSTNLYGGGIIYSSGA 886  
Qy 847 VTFSCSGTCQFSGNOADINPNQSLSNVQGAIVAKTSLISGSDAGTSYIFSGNSVST 906  
Db 887 VTLNLSGTFGIGTNSVINTATSQDA-DIQGGGIYATTSLSI--NQCNPILFSSNSAAT 943  
Qy 907 GKSQTTQIAGGAIYSPVTL---NCPATFSNNTASIAATPKTSSDGGSGNSIKDTIGGA 963  
Db 944 KKTSTTKIAGGAIFAATVIENNSQPIIFLNSAK-----SEATTAATAGN-KDSCGGA 997  
Qy 964 IAGTAITLSGVSF--PSGNTADLGAIG--TLANANTPSATSGSQNSITEKITLENPSFI 1019  
Db 998 IAANSVTLTNPNPEITFKGNAETGAIGCIDLTNGSPPRKVS-----IADNGSVL 1047

Qy 1020 FERNOA-NKRGAIYSPSYSIKNNITFENONTSTHDGSAIYFTKDATIESLGSVLFTGNV 1078  
Db 1048 FODNALNKGGAIIYETDTSRTGATFIGNSKHDSAICCSTALTALPNSQIIFENNV 1107  
Qy 1079 TATQASSATSGONTANYGAAIFGDPGCTQSQTDAITLLASSGNITFSNNLSQNOG 1138  
Db 1108 TETTATTKAS-----INNLAGAIYNNET-----SDVTISLSAENGSIFFKNLC----- 1152  
Qy 1139 DTPASKFCSIAGYKL-SLOAAKGGKTSFFDCVHTSTTKTGSTQNVYETDILNKEENSNP 1197  
Db 1153 -TATNKYCSIAGNVFTAEASAGAKAISFYDAVNSTKETNA-----QELKNEKATS-- 1204  
Qy 1198 YGTTVFVSELSHENKSYIPONAILHNGTLVLKREKTELHVVSFEQKESGLIMEPQAVLSN 1257  
Db 1205 -TGTTLFSGELHENKSYIPQKVTFAGHNLILCKNAELSVVFTQSPGTTITMPCGSVLSN 1263  
Qy 1258 QNIANGALAINGLTIDLSSMGTPQAGEIFSPPELRIVATSSASGGSGVSSSIPTNPKRI 1317  
Db 1264 HSKAEGGAIINNVIIDEFSEIVPTKDNATVAPPTLKVSRTNAD----- 1306  
Qy 1318 SAAVPSGSNAITPTMSENKVFITGDLTLIDPNGNFYQNPMLGSLDLDVPLIKLPTNTSD-V 1376  
Db 1307 -----SKDKIDITGTVTLLDPNGNLYQNSYLGEDRDTITLFDNDSASGAV 1351  
Qy 1377 QVYDLTSLGDLFPQGYMGTTWLTDSNPQTGKLOARWTFDYRRWVYIPRDNHFIANSILG 1436  
Db 1352 TATNVTLOGNLCAKGYLGTWNLDPNSSGSKILKWTFDKYLRWPIPRDNHFIINSWG 1411  
Qy 1437 SONSIMIVKQGLINMLNAREDDIAYNNFVSGVGTFLAQOQCTPLSEFYSYSGTSVA 1496  
Db 1412 AQNSLTVTKQGLGNLNNAREDPAFNFWASAIQSFLRKEVSRNSDSFTYHGRGYTAA 1471  
Qy 1497 IDAKPRQDFILCAAPSKIVGKTKAIKMMHNYPHKSEYSAQSVYCGKPYFLNKKQHSW 1556  
Db 1472 VDAKPRQDFILCAAPSVQFHAESYHLDNYKHGSGHSTQASLSYAGNIFYFPATRSR-- 1529  
Qy 1557 ALPFLIQGVVSYGHKHKHDTTLLYPSIHERNKGWDELGLADLRISMDLKPEKSKSSKRI 1616  
Db 1530 --PILFQGVATYGYWQHDTTYTPSEENKMANWDSIAWFLDLRFSVDLKEPQHPSTARL 1587  
Qy 1617 TVYGELEYSSIRKQFTEIDYDPRHFDCCAYRNLSLPVGCAGEGAIMNCNIMLYNKALAA 1676  
Db 1588 TPTTEAEVTRIRQKFTELDYDPRGSFACSIGNLAIPTGFSVDGALAMREIILYNNKVSAA 1647  
Qy 1677 YMPSIYRNPNVCKYRVLSSNEAGQVTCGVPTTSARAEBYSTQILYGLPFWTLXGNTIDVG 1736  
Db 1648 YLPVILRNPNKATYEVLTSTKEGNNVNVNLPTRNAARAESSQIYLGSYWTLTYTIDAS 1707  
Qy 1737 MYTLSQMTSCGARMIF 1752  
Db 1708 MNTLVQMANGIRFVF 1723

## RESULT 6

HB6557  
polymorphic membrane protein B Family [imported] - Chlamydomophila pneumoniae (strain J  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #texC\_change 02-Mar-2001  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: HB6557  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1723 <STO>  
A:Cross-references: GB:BA000008; NID:g8978911; PIDN:BAA98746.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_20

Query Match			
Best Local Similarity 34.8%; Score 3111.5; DB 2; Length 1723;			
Matches 727; Conservative 284; Mismatches 528; Indels 397; Gaps 45;			
QY	1	MKWLATATFAAALVPSVSGFCPEPEKELNFSRVETSSSTT-----FTETIGEA-CAEYI	53
DB	1	MKWLATATFAAALVPSVSGFCPEPEKELNFSRVETSSSTT-----FTETIGEA-CAEYI	60
QY	54	VSNASFTFTNIPPTDTPPTNSNS-----SSSGET-----	86
DB	61	IVGDITFSTFTNIPVPTPDANDSSNSKSGSSSGATSLIRSNLHSDPFTKDSVL	120
QY	87	-----ASVSDSDSTTTTPDPKGGGAFY-NAHSG	114
DB	121	DLYLHLPFASNTLNPALLSSSSSGSSSSSSSGSASAVVAADPKGGAAYNEANG	180
QY	115	VLSFMTFSGTESLTLSEIKMTGEGAIKFSOGELLTDLTSLTIONNLSGGAIF---	171
DB	181	TLFTTDSGNPGLTLONLKMTGDGAAIYSKGLPVLFTGLKLTFTGNESOKSGGAAYTEG	240
QY	172	-----GG-----STIS	177
DB	241	ALTTQAIQEAFTGTNTSAGOGAIYVKEATLFLNALDSLKFEKNTSGQAGGGIYTESTLT	300
QY	178	LSGITKA-TFSCNSAEVAPVKKPTPEKAQTASETSKSSSSSGNSGDSVSPSSRAEPAA	236
DB	301	ISNITKISIEFISNKASVPAPAPPTSP-----APSLINSTTIDTSTLOTTRAASATPAVA	355
QY	237	NLOSHFICATATPAATDTETSTPSHKPGSGGAIYAKGLTI-----	278
DB	356	PV-----AAVTPPTISTOETA-----GNGGAIYAKOGISISTFTKDLTFKNSASVDAT	403
QY	279	-----ADSOEV-----LFSINKATDGGAIKAEKDVSEFENITSLKV	314
DB	404	LFVDSSTIGESGAIFAADSIQIOCTGTLTFSGNTANKSGGGIYAVGVTTLEDIANLKM	463
QY	315	QTNGABEKGAIYAKGDLSTQSSK-QSLFNSNYSKOGGALYVEGGINFQDLEEFIRKN	373
DB	464	TNNTCKGEGAIYTKKALTINNGAILTTESGNTSTDNGGAIFAVGGITLSDLVEVFRSKN	523
QY	374	KAGTFETKTLTSLKAQASAGNADAWASSPOSGGATTVSDSGSSGSDSDTSETVP	433
DB	524	KTGNV-----SAPITKA-----ASNTAPVWSSSTTAASPAV-----PAAAAAP	561
QY	434	VT--AKGGGLYTDKNLSITNITGIIIEIANNKATDVGGAIVKGLTLCENSHRLOFTKNSS	491
DB	562	VYNAAGGALYSTEGTLVSGITSLIFENNECONQGGAYVYKTFQCSDSHRLQFTSNKA	621
QY	492	DKOGGIYGEDNTLNLTKTLFQNTAKEGGGLFIKGTDKALMTGLDSCFLINNTS	551
DB	622	ADBGGLYCGDDVTLNLTKTLFQNSSEKKGGLSL-ASGKSLTWTLSLSPCLNANTA	680
QY	552	EXHGGGAFYTKELSQYTT---SDVEITPGITPVHGETVITGNKSTGGNGGVCCTKRLAS	608
DB	681	KENGGGAPENIVLFTTPTTPNEPAPVQOQPVYGBALVTGNTAT-KSGGGIYTKNAAFS	739
QY	609	NQISISIGNSAAENGGAHTCPDSPPTADTAEOPAAAAAATSTPKSAPYSTALSTPSSS	668
DB	740	NLSVTFDQNTSSENG-----ALLTQKAAAD-----	765
QY	669	TVSSLLTAAASQASPATSNKETQDPNADPTDLLIDYVVDPTTISKNTAK-KGGGIYAKKAK	727
DB	766	-----NTDCSFYITVNTINNTATNGGGGIAGGKAH	797
QY	728	MSRIDQLNISENSATBIGGICCKESLELDALVSLVTENLVGKEGGGLHAKVTNINSLK	787
DB	798	FDRIDNLTVQSNQAKK-GGGVTLDELALILEKVTIGVSQNTATFESGGGIYAKDIQALP	856
QY	788	SGFSFNNKANSSTGVATTASAPAAAAAASLQAAAAAASPSPATPYSGVVGGAIGE-K	846
DB	857	GSFTITDNKVETSLT-----TSTNLYGGGGIYSSGA	886

QY	847	VTFQSCSGTCQFSGNOAIDNNPSSQSLNVOGGAIYAKTSLISGSSDAGTGYIFSGNSVST	906
DB	887	VTLTNSGTGFGTIGNSVINTATSQDA-DIOGGGIYATTSLSI--NQCNTPIILFNSNSAAT	943
QY	907	GKSQTTGOTIAGGAIYSPTVTL---NCPATFSNNATATPKTSSDGGSSSSNKTIDTIGGA	963
DB	944	KKTSTTKQTAGGAIFAASAVTIENNSQPIIFLNSAK-----SEATTATAGN--KDS:GGA	997
QY	964	IAGTATLSCVSR--FSGNTADLGAAG--TLANANTPSATSGSQNSITEKITLENCSFI	1019
DB	998	IAANSVLTNNPEITKGNVETGGAIGCIDLTLNGSPPKVS-----TADNUSVL	1047
QY	1020	FERNOA-NRGAISYSPSVSIKGNITFNQNTSTHDSAIYFTKDATIESLGSVLTGNV	1078
DB	1048	FQNSALNRGAISYGETIDISRTGATFIGNSKHDSGAICCSALTALAPNSQLIFEHNKV	1107
QY	1079	TATQASSATSGQNTANYCAAIFGPGTTOSSOTDAILTLASSGNITFSSNSLOINOG	1138
DB	1108	TETATTAKS-----INNLGAAYGNNET-----SDTISLSAENGSIFFKNLNC-----	1152
QY	1139	DTPASKFCSTIAGYVKL-SLQAAKGTISFPDCVHTSTKKTGSTQNVVETLDINKEEHNP	1197
DB	1153	-TATNYCSTIAGNVKTAIEASAGKALSFYDANVSTKETNA-----QELKLNKAT'S--	1204
QY	1198	YTGTVFSSSELHENKSYIPQNAIILHNGTLVKEKTELHVVSFEQEGSKLIMEPGAVLSN	1257
DB	1205	-TGTLFSGELHENKSYIPQVTFAGHNLTLGKNAELSVVSFTQSPGTTITMGPGSVLSN	1263
QY	1258	ONTANGALAINGLTIDLSSMGTPOAGEIFSPPELRIVATTVASSGSGSVSSSPTNPKRI	1317
DB	1264	HSKEAGGIALNNYIIDFSEIVPTKDNATVAPPTLKLVSRTNAD-----	1306
QY	1318	SAAPVSGSAATPTMSSENKVELGDLTLIDPNGNFYONPMLGSDLDVPLIKLPTNTSD-V	1376
DB	1307	-----SKDKIDITGTVTLDPNGLNQLYONSLYGEDRDLTLFNIDNSAGAV	1351
QY	1377	QYVDLTLSGLDFPKQYMGMTWLTDSNPQTQKQARWTFDYRRWVYIPRONHFFVANSILG	1436
DB	1352	TATNVTLOGNLGAKGKVLGTWNLDPNSSGSKIIILKWTFDKYLWVYIPRONHFFVINSIG	1411
QY	1437	SQNSMIVVQGLNLMNNAFDDIAYNNFWVGVTFFLAQOQTPLSEEFYSYRGTSTVA	1496
DB	1412	AQNSLVTVKGIGLNNLNAFDDIAYNNFWVGVTFFLAQOQTPLSEEFYSYRGTSTVA	1471
QY	1497	IDAKPQDFTLGAFAKIVGKTKAIKMMHNYFHKGSYSYQASVYGGKFLYLLNCKHGW	1556
DB	1472	VDAPKQEFILGAFAKIVGKTKAIKMMHNYFHKGSYSYQASVYGGKFLYLLNCKHGW	1529
QY	1557	ALPFLIQGVVSYGHIKHTTTLTLYPSIHERNKGDWEDLGLWADLRISMOLKEPSKDSKRI	1616
DB	1530	--PILFQGVATYGMQHDITTYYPSTIEEKNANWDSIAWFLDLRFSVDLKEPQPHSTARL	1587
QY	1617	TVYGELEYSSIROKQTEIDYDPRHEDDCAYRNLSLPVGCAGEAIMNCNLMYKALALA	1676
DB	1588	TFYTEAEYTRIOKEKTELDYDPRFSACSAGYLAIPGTFSDVGLAWREILLYNKVSAA	1647
QY	1677	YMSIYRNNPVCKRYLLSSNEAGQVCGVPTTSARAEYSTQLYLPFWTLGYNTYIDVG	1736
DB	1648	YLPVILRNNPKATYEVLSLTKGKGNVNVLPTRNAAREVSSQIYLSYWTLYGTYTTIDAS	1707
QY	1737	MYTLSQMTSCGARMIF	1752
DB	1708	MNTLVQMANGIRFVF	1723

RESULT 7

C81601

polymorphic membrane protein B/C family CP0212 [imported] - Chlamydia pneumoniae  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: C81601  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg







Db	487	AGGAIYFETEDFTLTGSTGVFTSTNTA-----KTAGALHTKGN-----TSFT	529
Qy	971	LSGVRESFGNTADLGAAGITFLANANPPSATSGNSQSI-----TEKITL-ENGSGFI	1019
Db	530	KNKALVPSGNSA-----TATATTTTDOECGGAILCNISEDIAPKSLTLTENEELS	581
Qy	1020	PERNOAKNR-GAIYSPSVSIKGN-NTTFNQNTSTHDDGSAIYPTKDATIESGSLVFTGNN	1077
Db	582	FINNTAKRSGGVIAPKCVISGESINFQDNTAETSGGAIY-SKNLSITANGPVSFTNN-	639
Qy	1078	VTATQASSATSGONTANTYNGAAIFGDCPCTQSSQTDAILTLASSGNITFTSNNSLGNQ	1137
Db	640	-----SGGKG-----GAIYIADSGE-----LSLEADGDTITSGN--RATE	673
Qy	1138	GDPASPASFCSIAGYVKVLSLQAARGKTIISFDCVHTSTKKTGSTQNVYETLQIN-----	1190
Db	674	GTSTPNSIHIGACAKYTKLAAAPGHTIVFYDDITWEPASAGCT-----IPEIYNDPVKAIY	730

QY	1191	---KEEN-----SNPYGTGIVFSS-ELEHNKSYIPQNAI-----LHNGT	1231
Db	731	PPQPKNPIASVPVVPVAPANTCTIVFSSGKLPSQDASIPANTTTILNOKINLAGGN	790
QY	1226	LVLKEKTELHVSEFEQESKLIMEGAVL--SNQIANGALAINGLIDILSSMGTPQAG	1283
Db	791	VLKEGATLQVSTFOQPOSTVFMDAGTTTLETTTNTNTDGSIDKLNLSVIALDQGR--	848
QY	1284	EIFSPEELRIVATTSASGSGSVSSSIPTNPKRISAAVPVSGSAATPTMTSENKVFLLGDL	1343
Db	849	-----MTIIVNTSGGLKIS-----GDL	867
QY	1344	TLIDPNGNYQNMLGSDLDVPLIKLPTNTSDVQVYDLT--LSGDLFPQKGYMGTWTLDLS	1401
Db	868	KFHNNESFYDNFGLKANLNLPFLDLSSTSGTVNLDDFNPISSMAAPDYGYQGSWTL--	926
QY	1402	NPQTG-----KLAQRTFTDYR-----RWVYIPRONHFYANSILGSONSMIVVQGLINN	1451
Db	926	VPKYVAGGKVTLVAEQALGYTPKPELRATLVP-----NSLWNAYVNIHSIQEIA--	976
QY	1452	MLNNAREDDIAYNFWVSGVGTFLAQOQPLSEFYSYRGTSVAIDAKPQDFTLGAAF	1511
Db	977	---TAMSDAPSHPCIWIGIGNAFHQDKOKENAGFPLISRGYIVGGSMTTQEVTFVAVP	1033
QY	1512	SKIVGKTK--AIKKMHNYFHKGS---EYSQASVYGG---KFLYFLLNKHQHGWALPFLIQ	1563
Db	1034	SOLFSGKSDYVSDIKSOVYAGSLCAQSSVVIPLHSSLRRHVSFLVLPFLPG-ETPLVLH	1092
QY	1564	GVYSYGHIKHDTTTLTPSITHRNKGDWEDLGLWADLRISMOLKEPSKDSKRITVYG---	1620
Db	1093	GOVSYGRHNHMTTKLAN-NTQCKSDWDSHFAVEVGGSL-----PVDLUNYRYLTSYSYV	1147
QY	1621	ELEYSSIRQKQFTEIDYDRPHFDCCAYRNLSLPVGCACV--EGATMNCNLMYKNLALAYM	1678
Db	1148	KLQVSVNQVQGFQEVAAADPRIFDASHLVNWSIPLMGLTTFKHESAKPPSALL-----	1203
QY	1679	PSIYRNPNVCYKRVLSNENAGOVICVPTRTSARAEYSTQLYGLPFWLYGNYTIDV3MY	1738
Db	1204	VDAYRDHPHC-LTSLTNGTISWSTFATNLSRQAFEAESGHLKLLHGLDCFASGSCEL3SS	1262
QY	1739	TLSQMTSCGARMIF	1752
Db	1263	SRSYNANCGTRYSF	1276
RESULT	11		
	A86611		
	probable outer membrane protein D family [imported] - Chlamydomophila pneumoniae		
	C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae		
	C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001		
	C:Accession: A86611		
	R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; et al.		
	Nucleic Acids Res. 28, 2311-2314, 2000		
	A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.		
	A:Reference number: A86491; MUID:20330349		

RESULT 11  
A86611  
probable outer membrane protein D family [imported] - Chlamydomydia pneumoniae  
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: A86611  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; S  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349

A:Accession: A86611  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1609 <STO>  
A:Cross-references: GB:BA000008; NID:g8979337; PIDN:BAA99171.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_21

```
Query Match      8.38; Score 743.5; DB 2; Length 1609;
Best Local Similarity 22.3%; Pred. No. 1.4e-21;
Matches 406; Conservative 265; Mismatches 687; Indels 459; Gaps 79;

Qy 165 LSGGAIFGSGTISLS-----GTTKATFSCNSA---EVPAPVKKPTEP--KAOTASE----- 210
Db 23 LSAGIAFAHLSHSELDLGVFNKQFHEHSAHVEAQTSLVKGSDPVNPSQKSEKVLVT 82
Qy 211 ----TSGSSSSS-----GND-----SVSSPSSSRAEP 233
Db 83 QVPLTQSGSGESLDLADANFLEHFOHLFEETTFVGIDQKLWMSDLDRNFSQPTQ---EP 139
Qy 234 AAANLQSHFICATAPAAQTOTETSTPSHKEG-----SGAIYAKGDLTIADSQEVLFSIN 289
Db 140 DTSNAVSEKI--SSDTKENRKDLETEDPSKKSGLREVSDLPKSPETAFAVAIASEDLEISEN 198
Qy 290 KATKD---GGAIF-----AEKDSFENI-----TSLKVOTNGAEKGGAIYAK 329
Db 199 ISARDPLOGLAFFKNTYSQISSEKSSFGIIFSGGANSGLGFENLKAPKSGAAVTSYD 258
Qy 330 GDLSIQSSKQSL-FNSNYSKGGALYVEGGINF-----ODLEEIRI-- 370
Db 259 RDIVFENLVKGLSFISCESLEDGSA---AGVNIIVTHCGDVTLTDCATGLDLEALRLVK 314
Qy 371 KYNKAGTETETK-----ITPLSKAQASAGNADAWASSPQSGSGATTVSDSGDSSG 423
Db 315 DFSRGGAVFTARNHEVQNNLHAGGILSVVGNKGAIVVEKNSAEKSGAGFACGFSFYVSNNE 374
Qy 424 SDSDTSEVPVYAKGGGLYTKNLISITNITGIIETANKA-----TD-VGGGAY- 471
Db 375 NYALWKENQALS--GGAISSASDIDIQNCSAIEFSGNQSILALGEHMLGLDFVGGGALA 432
Qy 472 VKGTLTCSNHRLOFLKNSDKQGGIYGEDNITLSNLTGKTLFQENTAKBEGGGLFTKG 531
Db 433 AOGTLTLRNNAVQCVCKNTSKTHGGAILA-GTVDLNETISEVAEKQNTAALTGGAL--SA 489
Qy 532 TDKALTWTGLDSFCLINTNTSEKHG---GAFVTKETISQTTSDVETIPGTPVHGETVI 587
Db 490 NDKVLIANNFGEILFEQNEVRNHHGAIYCGCRSNPKLQKQDSGENINIGNS---GATTF 546
Qy 588 TGNKST-----GGNGGGVCTKRLAL-SNLQSIISGNSAAEN-----GGGAHT 629
Db 547 LKNKASVLEVTQADYAGGALWGHVNLDSNSGNIQFIGNIGSTFWIGEVYGGGAIL 606
Qy 630 CPDSPTTADTA-----EQPAASAAATSPKSPAPVSTALSTPSSSTVSSLTLLAASSQ 681
Db 607 STDRTVTISNNSGDVYFKNGCQCLAQKYVAPQETAPVEDASS--TNKDEKSLNACSHGDH 665
Qy 682 ASPATSNKETQDPNADTDLDDYVDVTITISKNATKKGGIYAKKAKMSRIDQLNISENSA 741
Db 666 YPPKTFEE-----VPPSLLEEHPV---VSSDTIRGGGAILAQ-----HIFITDNTG 709
Qy 742 TEIGGGICCKESLELDALVLSLVTEINLVGKEGGGLHAKTV-NISNLKSGFSFNKAN-K-S 799
Db 710 -----NLRFSGNL-----GGGEESTVGDALVGGGALLSTNEVNC 746
Qy 800 SSTGVATTTASAPAAAAAALQAAAAAAPPSPATPTYGVVGGAIYGEKVTFFSQCSGTCQFS 859
Db 747 SNQNVFSDNVTNSGCDs-----GGAILAKKV-----DIS 776
Qy 860 GNOAID--NNFSQSLNVQGGAIYAKTSLSTGSSDAGTSTYFSGNSVSTGKSQTTGQTAG 917
Db 777 ANHSVEFYVSGSGKRF---GGAVCALNE-SVNITD-----NGSAVSFSKNRT--RLGG 822
```

```
Qy 918 GAIYSPVTTLNCPATFSNNTASIAIATPKT---SSBDGSGNSIKDTIGGAIAGTAITL--- 971
Db 823 AGVAAPOGVS---TICGNQCNIAFKENFVFGSENQSRGG-----GAIANSVSVNTQDN 872
Qy 972 -SGVSRFSGNTADLGAAI--GTLANANTPSATGSSQNSIITEKITLNGSFIFERNQANKR 1028
Db 873 AGDILFVSNSTGSGGAIFVGSV-----ASEGS-NPRTLITGNSGDILFAKNSTQTA 925
Qy 1029 GAIYSPSVSIKNNITFNQNTSTHDGSAIYFTKDATIESLGSVLTGNNVTATQASATS 1088
Db 926 ASL-----SEKDSFGGAIYQNLKIVKNAGNVSYGNR-----APS 962
Qy 1089 GQNTNTANYGAAIFGDPGTTQSSQTDAILTLASSGNITFNNSLQNNQGTDPASKFCSI 1148
Db 963 GAGVOIA-----DGGT-----VCLAEAGGDILFEGNI--NFDGSEFNAIHLGCGN 1003
Qy 1149 AGYKVLQQAOKGTISFFDCVHTSTKTKGTQNYE-----TLDIN---KEENSN 1196
Db 1004 DSKI-VELSAVQDKNITFQDAI---TVEENTIRGLPKDKVSPLSAPSLIFNSKPKQDOSAQ 1059
Qy 1197 PYTGTVFSSSELHENKSVIPQNAIILHNGTILVLKTELVHVVSFEQKESKLIMEPGAVL- 1255
Db 1060 HHEGTIRFS---RGVSKIPQIAAQEGTTLALSQNAELWLAGLKOETGSSIVLSAGSTLR 1115
Qy 1256 -----SNQNIANGALAIN-----CLTIDLSMG 1278
Db 1116 IFDSQVDSSAPLPTEENKEETLVSAQVQINMSSPTPNKDKAYDTPVLADIISITVDLSSFV 1175
Qy 1279 TPQAGEIFSPPELRIVATTSSASGGSVSSSIPTNPKRISAAVSPGSAATPTTSENKVF 1338
Db 1176 PEQDGTLPPLPEI-----IIPKGTKLHNSAI----- 1201
Qy 1339 LTGDLTLIDPNGNYQN-PMLGSDLDVPLIKLPT-----NTSDVQVYDVLTLISG 1385
Db 1202 ---DLKIIDPTNVGYENHALLSSHKDIPLISLKAETAGMTGPTADASLSNIKI-DVSLPS 1257
Qy 1386 DLFPQKVMGTWTLDNSNPOTGKLOARWTFDTPYRRWYTPRONHFEVANSILGSONSMIVVK 1445
Db 1358 ITPATYGHGTWMS-ESKMEDGRLVVGWQPTYK--LNPEKQAGALVNLNLWHSYTDLRALK 1314
Qy 1446 QGLINNMNNARFDIAYNNFWVSGVGTFFLAQOGTPISEEFYSYRSRGTSAIDAKPRODF 1505
Db 1315 QEIFAHTHTIAQRMELDFSTNVWVGSLGVEDQNTIGEDGFKHHLTYALGLDTQLVEDF 1374
Qy 1506 ILGAAPSKIVCKTKAKIMHNYFHKGSEYSYQASVYGKFLYLLNKHQHALPFLIOGV 1565
Db 1375 LIGGCFQOFFKTES---QSYKAKNDVKSYMGAAYAGIL-----AGPWLKGA 1419
Qy 1566 VSYGHIKHDTTTLTPSIHERNKGWEDLGWLA---DLRISMDLKEPSKOSKRTIVYGE 1621
Db 1420 FVYGNINNDLTDYGLT-GISTGSGWIGKGFAGTYSIDYRYIVNPRFISAIVSTVVPVE 1478
Qy 1622 LEYSIRQKQFTEIDYDPRHDDCAYRNLSLPVGCAGEGAIMNCNILMYNKLALAYMPSI 1681
Db 1479 AEYVRIDLPEISEQKEVTRFQKTRFENVAIPFGFALEHAYSRGSAEVNSVOLAYPVDV 1538
Qy 1682 YRNPV---CKYRVLSSNEAG-QVICGV-PTRTSARAEYSTQLYLGPFWTLGNYTTIDV 1735
Db 1539 YRKGPSVLITLUDAAYSWKSYGVDPICKAWKARLSNNTENWS--YLSYTLAFNFEWREDL 1596
Qy 1736 GMYTLTQMTSCGARMIF 1752
Db 1597 IAYDFNG-----GIRLIIF 1609
```

RESULT 12

H72013

polymorphic membrane protein D family CP0897 [imported] - Chlamydothila pneumoniae (S  
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: H72013; F81526  
R:Kalmann, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

[illegible]

QY 1736 GMYTSLQWTCGARMIF 1752  
| : : : : :  
Db 1597 IAYDFNG---GIRIIF 1609

RESULT 13  
H71468  
probable outer membrane protein D - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: H71468  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
A:Reference number: A71570; MUID:99000809  
A:Accession: H71468  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1531 <ARN>  
A:Cross-references: GB:AE001353; GB:AE001273; NID:g3329271; PIDN:AAC68408.1; PID:g332927  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: pmppD

Query Match 7.5%, Score 674; DB 2; Length 1531;  
Best Local Similarity 21.7%; Pred. No. 2.8e-24;  
Matches 419; Conservative 251; Mismatches 659; Indels 502; Gaps 86;

QY 8 AVFAALVPSVSGFC-----PPKELNFSRVETSSSTFTT 42  
| : : : : :  
Db 17 SWAAIILASVGLASCVDLHAGQSVNELVYVGPQAVLLDQIRDLFVSGKDSQAEGYR 76

QY 43 ETICEAGAEYIVSGNASFTKTNPTTDTTPT-----NSNSSSSGGETAS-VSEDSDS 95  
| : : : : :  
Db 77 LIVGDPSSFOBKADATLPKGVQEQTLFSVTPVVGQVDDQVSSQGLCSFTSSNLS 136

QY 96 TTTTPDPKGGAFYNAHSGVLSFMRGTEGLTLSEIKMTGEGGAIFSOCELLFTDLTS 155  
| : : : : :  
Db 137 -----PRDGESFLG-----IAFVGDSKAG-ITLTDVKASLSGAALYSTEDLIF----- 179

QY 156 LTIONNLSQLSGGAIFGG-STPISLGGITKA-----TFCNSAEVPAVKKKTEPKAQTASE 210  
| : : : : :  
Db 180 -----EKIGGLEFASCSLSQGGACAAQSIILHDCQGLQV-----KHCTTAVN 223

QY 211 TSGSSSSS-----GNDVSSPSSRAEPAANLQSHFICATATPAQTDTETSTSHKP-- 264  
| : : : : :  
Db 224 AEGSSANDHLFGGGAFFVTGSLSGE-----KSLYPAGDMVNVVANDGALSFEKGSANF 277

QY 265 GSGGAIYAKGD-LTIADSQEVLFSINKATKDGGAIFAEKDVSFENITSLKVQTN---GAE 320  
| : : : : :  
Db 278 ANGGAIAASGRVLFVANDKKTSTFIENRAL-SGGAIAASSDIAFQNCALFPGKNCALGTE 336

QY 321 EK-----GGAIYAKGDLSTQSKQSLFNSNYKQGGALY-----VEGGINFQD----- 364  
| : : : : :  
Db 337 DKSLGGGAISSLGLTVLLQGHGTCIDKNESASQGGAIKQCNQISDNQGVVFRDSTAC 396

QY 365 -----LEIRIKYNKAG-TFETKKTITLPSLKAQASAGNADAWASSPQSGSGATTVS 415  
| : : : : :  
Db 397 LGGGAIAAQEIVSTQNNQAGLSFEGGK-----ASFGGIACGFSFSSAGGA----- 441

QY 416 DSGSSSGSDSDTSETPVTAAGGGLYTDKNLSLTNITGITEIANNNKATDVGGAAYKGT 475  
| : : : : :  
Db 442 -----SVLGTIDISKN-----LGAISFSRTL 462

QY 476 LTCENSHRLQFLKNSSDKQGGIYGEDNITLSNLTKTLFOENTAKBEGGLFKIGTDKA 535  
| : : : : :  
Db 463 CTTSDLGQMEY-----QGGALFGENISLSNAGVLTPEKDNIVKT-----FASNGKI 509

QY 536 LTMGTGLDSCFLINNTSEKHGGGAFVTKETISQTYTSDVETIPGTPVHGTEVITGNKSTGG 595  
| : : : : :  
Db 510 L-----GGGAILAT-----GKVEITNNS----- 527

QY 596 NGGVCVCTKRLALSNLQISISGNSAAENGGAHTCPDFFPTADTAEQPAAASAATSPKS 655  
| : : : : :  
Db 528 -----EGISFTGNARA-----PQALPQOE---EFP-----LFSKKEG 556

QY 656 APVSTALSTPSSSTVSSLTLLAASSQASPATSNKETQDPNADTDL-----IDYVVDI 708  
| : : : : :  
Db 557 RPLSSYSGGGAILGREVAIL---HNAAVVEQNRQLOCSSEEAATLLGCCGGGAVHGMDS 613

QY 709 TISKNTAKKGGGIYAKKAKMSRIDQLNISENSATEIGGICKESLELDALVLSLVTE 768  
| : : : : :  
Db 614 SYVNSVVRGNNYA-----MGQVSGGALLSKTVQLAGNSVDFSRNI 657

QY 769 VKREGGLHAKTVNISNLKSGF-SFSNNKANSSSTGVATTASAPAAAAAQLQAAAAAPS 827  
| : : : : :  
Db 658 ASLGGGALQASEGNCELVDNGYVLFDRNRRG----- 688

QY 828 SPATPTSGVVVGAI---YGEKVTFSQCSTGCFSGNQAI----- 864  
| : : : : :  
Db 689 -----VYGGAIISCLRGD-VVISGNKGRVEFKDNIATRLVYVEETVEKVEVEPAEQ 738

QY 865 -DNN-----PSQSSLNVOGGAIYAKTSLSIGSSDAGTSVIFSGNSVSTGSKSQTGQIA 916  
| : : : : :  
Db 739 KDNNELSFLGRAEQSFITAAQALFA-----SEGDLS---PESSISSEELAKRRECA 788

QY 917 GGAIYSPVTTLNCPATFSNNTASITATPKTSEDSGSSGSIKDTIGGAIACTAITLSGVS 976  
| : : : : :  
Db 789 GGAIFAKRVRI-----VDNQEAUVFS-----NNFSDIYGGAI----- 820

QY 977 FSGNTADLGAAGTLANANTPSATSGSQNSITEKITILENGSFIFERNQANKRGAIYSPV 1036  
| : : : : :  
Db 821 FTG-----SLREEDKLD-----GQI---PEV 838

QY 1037 SIKGN--NITFNQNTS-----THDGSIAVFTKDATI-ESLGSVLTGNNVTAQASSAT 1087  
| : : : : :  
Db 839 LISGNAGDVFSGNSKRDEHLPHTGGAICTQNLITISQNTGNVLFY-NW-----AC 890

QY 1088 SQQNTNTANTGAALFGDPGTQSSQTDAILTLASSGNITFSNNSLQNNQOGDTPASKFCS 1147  
| : : : : :  
Db 891 SG-----GAVRIEDHGN-----VLLEAFGDDIVFKGNSFPRAQG-SDAIYFAG 932

QY 1148 IAGVVKLSLQAAGKKTISFFDCV---HTSKTKTGSTQNVYETLDINKEENSNPYTGIVF 1204  
| : : : : :  
Db 933 KESHI-TALNATEGHAIVFDHALVFENLEERKSA-----EVLINRENPQ-YTGSIRF 984

QY 1205 SSELHENKSYIPQNAIILHNGTLVLKTEHLVVSFEQEGKSLIMEPCAVLSNQNIANGA 1264  
| : : : : :  
Db 985 L-----EAESKVPQCIHVQOQSLELLNGATLCSYCFKODAGAKVLAAGAKLK----- 1032

QY 1265 LAINGLTIDLSMGTP-QAGEIFSPPELRIYVATTSSAGGS-----GVSSSIP-TNPKRI 1317  
| : : : : :  
Db 1033 -----ILDSCTPVQOQGHAIKPEAETESSEPEGAHSLWIAKNAQTTPVPMVDIHTI 1083

QY 1318 SAAVPS-GSAATPTTMSKNKVLFG-----DLTLIDPNGNFYON-PMLGSDLOVPL 1366  
| : : : : :  
Db 1084 SVDLASFSSSQOQETVEAPQVIVPGSVYRSGELNLELVNTGTGYNHALLKNEAKVPL 1143

QY 1367 IKLPT-----NTSDVQVYDLT--LSGDLFPQKGYMGWTWILDSNPQTKLOARW 1412  
| : : : : :  
Db 1144 MSFVASGDEASAEISNLVSVDLOIHVVTPETEEDTY---GHMGDWS-EAKIQDGTLVISW 1199

QY 1413 TFDYTRRWYIIPRDNHIFYANSILGSONSMIVV-----KQGLINNNLNARPDITAYN--- 1464  
| : : : : :  
Db 1200 NPTGYR-----LDPKAGALVFNALMEGAVLSALKNARF---AHNLTA 1240

QY 1465 -----NFWVSGVGTFF--LAQGGTPLSEFSYSRGTSVAIDAKPRODFILGAAFSK 1513  
| : : : : :  
Db 1241 QRMEDYSTNVWGFAGFGFRTLSAENLVAIDGYKAGYGASAGVDIQLMEDFVLGVSAA 1300

QY 1514 IVGTIKAIKMHNYFHKGSEYYSQAQSVYGGKFLYLLNKQHGHALPFLIQGWWSYGHK 1573  
| : : : : :  
Db 1301 FLGRMDSQKFDPAEYSRKG-----VVGSVYTGFL-----AGSNFFKGOYSLGETON 1345

QY 1574 DTTTLTYSIHERNKGDWEDLGLWADLRIS-MDLKEPSKDS--SKRITVYGELEYSSIRQK 1630

```
Db 1346 DMKTRYGVIGE--SSASWTSRGLADALVYRSLVGVVRPTFYALHFNPNPYEVSYSKMP 1404
Qy 1631 QFTEIDYDPRHFDCAAYRNLSPVGCAGVATMNCNLKALAYMPSIYRNPNVCKY 1690
Db 1405 GTEQGRARSFEDASLTNITPLGKMFELAFIKGFSEVNSLGISYAWEARKKVEGAV 1464
Qy 1691 RVLSSNEAGQVIGVPTRTSARAEXSTQLYLGPFWFLYGNITDV---GMYT-----LS 1741
Db 1465 QLL---EAGFDWEGAP-MDLPRQLRVALENTEWSSYFSTVLGLTAFCGGFTSTDSKLG 1520
Qy 1742 QMTSCGARMIF 1752
Db 1521 YEANTGLRLIF 1531

RESULT 14
D86546
polymorphic outer membrane protein G family [imported] - Chlamydophila pneumoniae (strain
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: D86546
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishi-
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86546
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <STO>
A:Cross-references: GB:BA000008; NID:g9978818; PIDN:BA98654.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: pmp_8

Query Match 7.4%; Score 665.5; DB 2; Length 930;
Best Local Similarity 25.0%; Pred. No. 3.4e-24;
Matches 292; Conservative 150; Mismatches 389; Indels 335; Gaps 48;

Qy 658 VSTALSTPSSSTVSSITLLAASSQASPATS-----NKETQDPNADTDLIDYVVD 707
Db 10 ISSTLVTP---ILLSTATVGADASLSPDTSDFCAGGSGTFTPKSTADANG-----TNYVLS 61
Qy 708 TTISKTAKKGGIYAKKAKMRIDQLNLSSENSATEIGGICCKESLELDALVLSLVTE 767
Db 62 GNVIINDACKG-----TALTG--CC-----FTET 83
Qy 768 LVGKEGGLHAKTVNLSLKSFSFNNKANSSTGCVATTASAPAAAAAALQAAAAAPS 827
Db 84 -----TGDLTFTGKYSFSFNTVDAGN-----AGAAAATTADKALTFT 122
Qy 828 SPATPYSGVGVGAIYGERKVTFSOC-----SGTCQFSGNOAIDNPNQSSSLNVQGGAI 880
Db 123 GFSNLSFIAAPGTVVASGKSTUSSAGALNTDGTILFSONVS-----NEANNNGGAI 175
Qy 881 YAKTISIGSDAGTSYIFSGNSVSTGKSQTTQIAGGAIYSPVTPLNCPATFSNNNTASI 940
Db 176 TAKT-LSI-----SGNTSSITFTSNAKKLGGAIYS-----SAAASISGNTGOL 218
Qy 941 ATPKTSSEDSGSGNSIKDTIGGAI---ACTAITLGSVRSFSGNTADLGAAGTLANANTP 997
Db 219 VPMNNKGETG-----GGALGFSEASSITQNSLFFSGNTA----- 253
Qy 998 SATSGQNSITEKITLENGSFIFERNQANKRGAIY-----SPSVSIKGN-NITFNQNTS 1050
Db 254 -----TDAAGKGAIYCEKGTGPTTTLTISGNKSLTFAENS 289
Qy 1051 THDGSIAIYTKDATIESLGSLVLTGNVNTATQASSATSGONTNTANYGAAIFGDPGTTQS 1110
Db 290 VTQGGAI-CAHGLDLSAAGPTLFSNNRC-----GNTAAGKGGAIALDSGS--- 334
Qy 1111 SOTDAILTLASSGNTTFSNNSLQNNQGDTPASKFCSTIAGYVKL-SLOAKAGKTISFFDC 1169
```

```
Db 335 -----LSLSANGODITFLGNTLTSTSAPTSTRNAYILGSSAKITNLRAAQGOSIYDP 388
Qy 1170 VHTSTKTKSTQNVYETLDINKENSNP--YGTGTIVFSSE---LHENK-----SVLPQN 1218
Db 389 IASNT--TGAS-----DVLTIINOPDSSPLDYSGTITVFSFGEKLSADAKAADNFTSILKOP 442
Qy 1219 AILHNGTILVLKKEKTELHVVSFEQEGSKLIMEPAGVLSNQNIANGALANGTLIDLSMGM 1278
Db 443 LALASGTALKGNVELDVNGFTQTEGSTLLMQPGTKLKAD---TEAISLTKLVVDLSAL- 498
Qy 1279 TPOAGEIFPPELRIVATTSSASGGSGVSSSIPTNPKRISAAVPSGAATPTPTMSENK-V 1337
Db 499 -----EGNKSVS-----IETAGANKTI 515
Qy 1338 FLTGDLTLDIPNGNFYQNPMLGSDLDVPLIKLPTNTSDVOVY--DLTSLGDL---FPCKGY 1393
Db 516 TLTSPVLFDQSSGNFYESTHTINOAFQPLVFTTAATAASDIYIDALLTSPVQTPPEHYGY 575
Qy 1394 MG-----TWTLDSPQTKLQARWTFDTY-----RRVVIYPRDNHFYAN-----SILG 1436
Db 576 QGHWEATWADTSTAKSGTM--TWVTTGYNPNPERRASVVP--DSLWASFTDITLQCIWT 631
Qy 1437 SQNSMTIVVQGLINNMNARNPDDIAYNFWVSGVGTFLAQOQTPLPSEEFYSYSGTISVA 1496
Db 632 SQANSIYQORGL-----WASGTANFFHKDKSGTNOAFRHKSYSYIVG 673
Qy 1497 IDAKPRQDFILGAAFSKIYVTKTKAIKKMHNYPHKSGSEYSYQASVY--GCKFLYFLLNKQH 1554
Db 674 GSAEDFSENI FSAFCQLPGKDKDLFIVENTSH-----NYLASLYLQHRAFLGGLPVP 728
Qy 1555 GWA-----LPFLIQGVWSYGHKKHDTTLYPSIHERNKGDWEDGLWGLADLRISMDLKEP 1608
Db 729 GSITDMLKDIPLILNAQLSYSTYKNDMDTRYTSYPEA-QGSWTNNSGALELGGSLALYLP 787
Qy 1609 SKDSSKRITVYGELEYSSI--RQKQTEIDYDPRHFDCAAYRNLSPVGCAGVGAJMN 1666
Db 788 -KEAPFGQYFPFLKFAVYSRQNEKESGAERAFDDGDLVNCSPVGIKRL-----K 840
Qy 1667 ILWYNK-----LALAYMPSIYRNPNVCKYRVLSSNEAGQVIGVPTRTSARAEXSTOLYLG 1722
Db 841 ISEDEKNPFISLAYIGDYRKNRPSRTSLMVSGASWTSCLKNLARQAFSLASGSHLTLS 900
Qy 1723 PWTLYGN-----YTIDVGM 1737
Db 901 PHVELSGEAYELRGSAAHIYNVDCGL 926

RESULT 15
A81591
polymorphic membrane protein G family Cp0307 [imported] - Chlamydophila pneumoniae (s
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81591
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38164.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0307

Query Match 7.4%; Score 665.5; DB 2; Length 930;
Best Local Similarity 25.0%; Pred. No. 3.4e-24;
Matches 292; Conservative 150; Mismatches 389; Indels 335; Gaps 48;
```

```
Qy 658 VSTALSTPSSSTVSSLLTLLAASSQASPATS-----NKETODPNADTDLIDYVD 707
Db 10 ISSTLVTP---ILLSIATYGADASLPTDSDGAGGSTFTPKSTADANG-----TNYVLS 61
Qy 708 TTISKNTAKGGGIYAKAKMSRIDQLNISENSATEIGGGICCKESLELDALVLSVTEN 767
Db 62 GNVYINDAGK-----TALTG---CC-----FTET 83
Qy 768 LVKREGGLHAKTVNISNLKSGFSNNKANSSTGVATTASAPAAAAAASLQAAAAAPS 827
Db 84 -----TGDLFTFGKGYSEFNTVDAGN-----AGAASTTADKALTFT 122
Qy 828 SPATPTYGVVGGAIYGEKVTFQC-----SGTCQFSGNOAIDNNPQSQSLNVQGGAI 880
Db 123 GFSNLSFTAAPGTTVASCKSTLSSAGALNLTGNTILFSONVS-----NEANNNGGAI 175
Qy 881 YAKTSLSIGSSDAGTSYIFSGNSYSTGKSQTTGQIAGGAIYPTVTLNCPATFNNNTASI 940
Db 176 TAKT-LSI-----SGNTSSITFTSNSAKKLGGAIIYS-----SAAASISGNTQQL 218
Qy 941 ATPKTSSEDSGSGNSIKDTIGGAI---AGTAITLGSVSRFSGNTADLGAATGTLANANTP 997
Db 219 VFMNKGEGT-----GALCFEASSITQNSLFFSGNTA----- 253
Qy 998 SATSGSONSITEKITLENGSFIERNQANKRGAIY-----SPSVSIKGN-NITFNQNTS 1050
Db 254 -----TDAAGKGAIYCEKTGETPTLTISGKSLTFAENSS 289
Qy 1051 THDGSAYFTKDATIESLGSVLFTGNVNTATQASSATSGQNTANTYGAALFGDPGTTOS 1110
Db 290 VTQGGAI-CAHGLDLSAAGPTLFSNNRC-----GNTAAGKGGAIAIADSGS--- 334
Qy 1111 SOTDAITLLASSGNITFNSNLSNONQODTPASKFCSTAGVVKL-SLQAAKGKTIISPPDC 1169
Db 335 -----LSLSANQGDITPLGNTLTSTAPTSTRNAILGSSAKITNLRAOQOSIYFYDP 388
Qy 1170 VHTSKTKGTQONVETLIDINKEENSNP--YTGTVFSE---LHENK-----SYIPQN 1218
Db 389 IASNT--TGAS---DVLTIINOPDSNPLDYSGTVFSGEKLSDAEAKAADNFTSILKQP 442
Qy 1219 AILHNGTILVKEKTELHVVSFEQEGSKLIMEPGAVLSNQNIANGALAINGITIDLSSMG 1278
Db 443 LALASGTLALKGNVELDVGFTQTEGSTLLMOPGPKLAD---TEATSLTKLVVDLSAL- 498
Qy 1279 TPQAGEIFSPPELRIVATTSSAGSGSVSSIPTNPKRISAAVPSGSAATPTMTSENK-V 1337
Db 499 -----EGNKSVS-----IETAGANKTI 515
Qy 1338 FLTGDLTLLIDPNFYQNPMLGSLDLVPLIKLPTNTSDVQVY-DLTLSGDL---FPQKGY 1393
Db 516 TLTSPLVFQDSSGNFYESHNTINQAFTPQVLVVTAATAASDIYIDALLTSPVOTPEPHYGY 575
Qy 1394 MG---TWTLDSNPQTGLQARWTFDTY-----RRWYIPRDNHFYAN-----SILG 1436
Db 576 QGHWEATWADTSTAKSGTM--TWVTGYNPNPERRASVVP--DSLWASFTDIRTLQOIMT 631
Qy 1437 SQNSMIVVKQGLINMLNLRNFDDIAYNNFWVSGVTFLAQOQTPLSEEFYSYRGTSVA 1496
Db 632 SQANSIYQORGL-----WASGTANEFHDKDSCGTNQAFPHKSYGYIVG 673
Qy 1497 IDAKPRODFILGAAFSKIVGKTKAIKMHNYFHKGEYSYQASVY--GCKFLYFLLNKQH 1554
Db 674 GSAEDFSENIFSVAPCQLFGKDKOLFIVENTSH-----NYLASLYLQHRAFUGGLPMPSPF 728
Qy 1555 GWA-----LPFLIOGVVSYGHKHDTTILYPSIHERNKGDWEDLGLWADLRISMDLKEP 1608
Db 729 GSITDMLKDIPLILNAOQUSYTTNDMDTTRYTSPEA--QGSWTNNSGALELGGSLALYLP 787
Qy 1609 SKDSKRTIVYGELEYSSI--RQKOFTEIDYDPRHFDQCAVRNLSLPVGCAGEAIMNCN 1666
Db 788 -KEAPFFQGYFFFLKFOAVYSRQONFESGAEAFADGDLVNCISIPVIRLE-----K 840
Qy 1667 ILMYNK-----LALAYMPSIYRNNPVCKYRVLSSNDAQVICGVPTRTSARAEYSTQLYLG 1722
```

```
Db 841 ISEDEKNNFEISLAYIGDVYRKNNPRSRSTLMVSGASWTSLCKNLARQAFLASAGSHLTL 900
Qy 1723 PFWTLYGN-----YTIDYGM 1737
Db 901 PHVELSGEAYELRGSAHIYNVDCGL 926
```

Search completed: October 25, 2001, 08:41:50  
Job time: 349 sec

EMBL:AE001361

ID AE001361 standard; DNA; PRO; 5601 BP.  
XX  
AC AE001361; AE001273;  
XX  
SV AE001361.1  
XX  
DT 22-JUL-1998 (Rel. 56, Created)  
DT 01-NOV-2000 (Rel. 65, Last updated, Version 5)  
XX  
DE Chlamydia trachomatis section 1 of 87 of the complete genome.  
XX  
KW .  
XX  
OS Chlamydia trachomatis  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
XX  
RN [1]  
RP 1-5601  
RX MEDLINE; 99000809.  
RX PUBMED; 9784136.  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia  
RT trachomatis";  
RL Science 282(5389):754-759 (1998).  
XX  
RN [2]  
RP 1-5601  
RX MEDLINE; 99206606.  
RX PUBMED; 10192388.  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
RL Nat. Genet. 21(4):385-389 (1999).  
XX  
RN [3]  
RP 1-5601  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;  
RT ;  
RL Submitted (20-MAY-1998) to the EMBL/GenBank/DDBJ databases.  
RL Program in Infectious Diseases, University of California, 235 Warren Hall,  
RL Berkeley, CA 94720-7360, USA  
XX  
DR SPTREMBL; 084881; 084881.  
DR SPTREMBL; 084882; 084882.  
DR SPTREMBL; 084883; 084883.  
XX  
FH Key Location/Qualifiers  
FH  
FT source 1..5601  
FT /db\_xref="taxon:813"  
FT /organism="Chlamydia trachomatis"  
FT /strain="D/UW-3/CX"  
FT CDS 323..640  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:084881"  
FT /transl\_table=11  
FT /gene="CT873"  
FT /product="hypothetical protein"  
FT /protein\_id="AAC68471.1"  
FT /translation="MSSNRMLATIKKITVLLLSKRKAGIRIDYCALALDAVEYSYLWFL  
FT VVGYAVYSSGVRWEAILEGETLSPGDRAQLDASLESAQVRLAGLMLDYWDGDSRLECKK  
FT I"  
FT CDS complement(815..3451)

```

FT      /codon_start=1
FT      /db_xref="SPTREMBL:O84882"
FT      /transl_table=11
FT      /gene="pmpI"
FT      /product="Putative Outer Membrane Protein I"
FT      /protein_id="AAC68472.1"
FT      /translation="MRPDHMFCCCLCAAILSSTAVLFGQDPLGETALLTKNPNHVVCTF
FT      FEDCTMESLFPALCAHASQDDPLYVLGNSYCWVSKLHITDPKEALFKEKGLDSIQNFR
FT      FLSFTDCSSKESSPSIIHQKNGQLSLRNGSMSFCRNHAEGSGGAI SADAFSLQHNLYLF
FT      TAFEENSSKGNNGGAIQAQTFSLSRNVSPISFARNRADLNGGAICCSNLICSGNVNPLFF
FT      TGNSATNGGAICCSIDLNTSEKGSLSLACNQETLFASNSAKEKGGAIYAKHMLRYNGP
FT      VSFINNSAKIGGAIAIQSGGSLSLAGEGSVLQNNQRTSDQGLVRNAIYLEKDAILS
FT      SLEARNGDILFFDPIVQESSKESPLPSSLQASVTSPTPATASPLVIQTSANRSVIFSS
FT      ERLSEEEKTPDNLTSQLOQPIELKSGRLVLKDRVLSAPSLSDQPQALLIMEAGTSLKT
FT      SSDLKLATLSIPLHSLDTEKSVTIHAPNLSIQKIFLSNSGDNFYENVLLSKEQNNIP
FT      LLTLSKEQSHLHLPDGNLSSHFGYQGDWTFWSKDSDEGHSLIANWTPKNYVPHPERQST
FT      LVANTLWNTYSDMQAVQSMINTIAHGGAYLFGTWGSAVSNLFYAHDSGKPIDNWHHRS
FT      LGYLFGISHTSLDDHSFCLAGQLLGKSSDSFITSTETTSYIATVQAQLATPLMKISAQ
FT      ACYNESIHELKTKYRSFSKEGFGSWHSVAVSGEVCASIPVSNNGSGLFSSFSIFSGLQG
FT      FSGTQDGFEESSGEIRSFSSASSFRNISLPMGITFEKKSQKTRNYYYFLGAYIQDLKRDV
FT      ESGPVVLLKNAVSWDAPMANLDSRAYMFRLTNQRALHRLQTLLNVSYVLRGQSHSYSLD
FT      LGTTYRF"
FT      CDS      3725..5500
FT      /codon_start=1
FT      /db_xref="SPTREMBL:O84883"
FT      /transl_table=11
FT      /gene="CT875"
FT      /product="hypothetical protein"
FT      /protein_id="AAC68473.1"
FT      /translation="MSIRGVGGNGNSRIPSHNGDGSNRRSQNTKGNNKVEDRVCSLYSS
FT      RSNENRESPYAVVDVSSMIESTPTSGETTRASRGVFSRFRQGLVRVADKVRRAVQCAWS
FT      SVSTRSSATRAAESGSSRTARGASSGYLREYSPSAARGLRMFTDFWRTRVLRQTSFM
FT      AGVFGNLDVNEARLMAAYTSECADHLEANKLAGPDGVAAAREIAKRWEQVRDLQDKGA
FT      ARKLLNDPLGRRTPNYQSKNPGEYTVGNMFMFYDGPQVANLQNVDTGFWLDMSNLSDVVL
FT      SREIQTGLRARATLEESMPMLENLEERFRRLQETCDAARTEIEESGWTRRESASRMEDGE
FT      AQGPSRAQQAQFSFVNECNSIEFSFGSFGEHVRVLCARVSRGLAAAGEAIRRCFSCKKG
FT      STHRYAPRDDLSPEGASLAETLARFADDMGIERGADGTYDIPLVDDWRRGVPSIEGEGS
FT      DSIYEIMMPIYEVMDMDLETRRSFAVQQGHYQDPRASDYDLPRASDYDLPRSPYPTPPL
FT      PPRYQLQNMDVEAGFREAVYASFVAGMNYVVTQPQERIPNSQQVEGILRDMLTNGSQT
FT      FRDLMRRWNREVDR"
FT
XX
SQ

```

Sequence 5601 BP; 1664 A; 1030 C; 1382 G; 1525 T; 0 other;

tctaggtgag	ctcgtgaagaa	gacgtattgt	aagaacagtt	tctataattt	ggtggtaatt	60
tttattttta	ataataaaaa	atacgcccta	attcagctaa	aacagccttg	tttttattaa	120
acccattttg	ttttctaacg	gaatcgaggt	gaatcatggt	ttgtccttta	cgcagttgct	180
gtactactca	taccagcatg	gagagcgtgt	ctttggaaga	gacctcgcct	cttttagaag	240
cgatatcaca	ggctgtgcaa	acatgcttaa	ccgtagccaa	gcacacagag	aacgctaaga	300
tttataatca	aaaacaaaaga	agatgtcatc	aaacaggatg	ctggcaacaa	ttaaaaaaat	360
tactgtgttg	ctgttgtcca	aaagaaaagc	tggtatcagg	atcgattact	gcgctctcgc	420
tctggatgca	gtcgaatatt	cctacttatg	gttttttagt	gtaggttatg	cggatatatg	480
ttcaggagta	cgttgggaag	cgatcttgga	aggagagact	ttatctcctg	gagatagagc	540
gcaactagat	gcctcttttag	agagtgtctca	agttaggtta	gctggtttga	tgttggacta	600
ttgggacgga	gattctcgcc	tagaatgcaa	gaagatataa	tcgccagggt	tcgagacaaa	660
gtcgggatct	gtgagcttac	agatgcggtt	gctctgaaac	acttctgtcg	gagaacttaa	720
atgtgatacg	cgacgtgaga	aaaaaatgaa	aagtcttaga	gagagatccc	taagactttt	780
ctcaagctct	aaaacgcaat	gctaggcaag	ggaactagaa	cctgtaagtg	gtccccagat	840
ccaggagta	actatggcct	tgcccgcgca	gtacgtaaga	cacatttaac	agcgtctgaa	900
gtctatgcag	agctctttga	ttcgtaagcc	tgaacatgta	ggctcgcgaa	tccaagttcg	960
ccataggagc	atcccaggag	acggcatttt	tgagtaacac	tacaggcccc	gattccacat	1020
cacgttttag	gtcttggtatg	taggctccca	gaaagtaata	gtagtttcgt	gttttttggg	1080
attttttttc	aaatgttatt	cccataggaa	gtgaaatatt	tctgaaagag	ctggcagaaa	1140
aggaccgaat	ctctcccga	ctctcctcaa	aaccgtcctg	tgttcctgaa	aatccttgca	1200
gttttagagaa	aatagagaag	gagctgaaca	gtccggaacc	attggataca	ataggaatcg	1260
atgcgcacac	ttctccgat	actgcaacgc	tatgccagga	tccgaatcct	tcttttagaga	1320
aggagcgata	ttttgttttt	agctcatgga	tactttcatt	atagcatgcc	tgtgcagaga	1380
ttttcattag	aggggtagcg	agttgcgctt	gtacagtagc	tatataggag	gtcgtttctg	1440
tagacgtaat	aaaggaatcg	gacgatttcc	cgagtaattg	tcctgcagcc	aagcagaaa	1500

aatggtcatc	taaactgtga	gtactgatac	cgaataggta	gccaaggcct	ctatgatgcc	1560
aattatcgat	aggtttccca	gagctgtcgt	gagcatagaa	taaattagaa	acagcagatc	1620
cccacgttcc	aaatagatag	gctcctccgt	gcgctattgt	attaatcabc	gactgcacag	1680
cttgcataac	ggaatagggtg	ttccaaagag	tggtcgcaac	gagtgtagat	tgacgttctg	1740
gatgaggcac	atagttttta	ggcggtccaa	tagcaatcag	agaatgccct	tcatacagaat	1800
ctttccaaga	aaaagtccaa	tctccttgat	atccaaagtg	agaagagagg	ttccccatcag	1860
gaagatgtaa	atgagattgc	tcttttagaga	gagtaaggag	aggaatattg	ttttgctctt	1920
tactgagaag	ctctacattt	tcataaaaaat	tctcatctcc	agaattagag	aggaagatct	1980
tttgatagaa	aagggttaggg	gcgtggatag	ttacgctttt	ttcagtatct	aaggaatgaa	2040
ggggaatact	tagcgtagct	aacttcaa	cagaggaagt	ttttaaagaa	gttcccgcct	2100
ccataatgag	gagagcttga	ggatcctgag	agagagaagg	cgcggaagg	acagcgcgat	2160
cttttaaaac	taagcgctcg	gatttccagt	cgataggctg	ctgtagttgg	gaagtgaggt	2220
tatcaggagt	tttttcttct	tcagaaagac	gttcgctcga	gaaaatcact	gaacggtttg	2280
cacttgtctg	aataactaaa	ggagatgcgg	tggtcggtgg	gggagaagtc	acgctggctt	2340
gcaaagagga	gggaagaggc	gattcttttg	tgctactttc	ttgtacaata	ggatcaaaaga	2400
aaagaatata	tccgttgcca	gcttctaagg	aagaaagaat	cgcatctttc	tctaagtaga	2460
tggtggtttt	tactagacct	tggtcgagg	tggttggtga	gttattctgg	aacagaacag	2520
atccttcacc	tgcaaggata	gagagactcc	ctccggactg	gatggcgata	gctccacctt	2580
ttttagcgct	gttggttaag	aaggaaacag	gaccgttata	acgcaatacc	atgtgcttgg	2640
cataaatagc	cccgcctttt	tcttttagcag	aattgcttgc	aaatagcggt	tcttggttac	2700
aagcaagaga	gagagagcct	ttttctgagg	tggttagatc	gctgatacaa	caaatagcgc	2760
ctccattcgt	ggcggagttt	ccagtgaata	agagagggtt	tacattccct	gaacaaataa	2820
gattactaca	gcaaataagc	ccgccatttt	aatccgcacg	attacggggc	aaagaaatag	2880
gcgacacatt	tctagataaa	gagaagggtt	gagcctgaat	ggctccgcca	tttcttttag	2940
aagaattctc	ttcaaaagct	gtgaaaagat	agttgtgttg	tagagaaaag	gcatccgcag	3000
agatggctcc	tccagagcct	tcagcatgat	ttcgacagaa	actcatgcta	ccattattgc	3060
gcaaggataa	ctgaccattc	ttttgatgaa	taatagaagg	agagctttcc	ttggaagagc	3120
aatctgtgaa	ggaaaaggag	cgaaaatttt	gaatggaaag	atctcctttt	tctttaaaaa	3180
gagcctcttt	gggttccgtg	atatggagtt	tagatacgaa	ccaacagtag	gaatttccaa	3240
gtacatacaa	aggatcatct	tgtgatgcac	gagcacaag	agcaggaaag	aggctctcca	3300
tggtacagtc	gtcaaaaaat	gtacagacga	catgattag	atttttagtg	aggagggcgg	3360
tttcacctaa	gggttccgtg	ccaaagagga	ccgctgtgga	tgacaaaata	gcagcacata	3420
gacaacagaa	gttcatatga	tcagggtcgca	tggtattccc	cctaagagtc	taaaagaaga	3480
ggttatgtca	aaaaaaagta	ataatttaca	gactttttcc	tcaagggttc	tgtttcatgt	3540
ctttcaagat	gaagagttaa	ggaagatttt	tggtattgtg	aattagaaaa	atcagaatta	3600
tcaaattaga	tagtttttaa	aaaatttaaa	aatcttttaa	gagatcttat	tttattatcg	3660
atttttttaa	ttttgatttt	taaaattaaa	ataagttttt	atttttagttt	tgggttggtt	3720
tgttatgagc	atcaggggag	taggaggcaa	cggaatagt	cgaatccctt	ctcataatgg	3780
ggatggatcg	aatcgagaa	gtcaaaaata	gaagggtaat	aataaagttg	aagatcgagt	3840
ttgttctcta	tattcatctc	gtagtaacga	aaatagagaa	tctccttatg	cagtagtaga	3900
cgtcagctct	atgatcgaga	gcaccccaac	gagtggagag	acgacaagag	cttcgctggg	3960
agtgttcagt	cgtttccaaa	gaggtttagt	acgagtagct	gacaaaagta	gacgagctgt	4020
tcagtgtgcg	tggtgttcag	tctctacaag	aagatcgctc	gcaacaagag	ccgcagaatc	4080
cggatcaagt	agtcgtactg	ctcgtggtgc	aagttctggg	tataggagag	attctccttc	4140
agcagctaga	gggtgtcgct	ttatgttcac	agatttctgg	agaactcggg	ttttacgcca	4200
gacctctcct	atggctggag	tttttgggaa	tcttgatgtg	aacgagggtc	gtttgatggc	4260
tgctgtacaa	agtgagtgcg	cggatcattt	agaagcgaac	aagttggctg	gccctgacgg	4320
ggtagcggcc	gcccgggaaa	ttgctaaaag	atgggagcaa	agagttagag	atctacaaga	4380
taaagggtgct	gcacgaaaat	tattaaatga	tccttttaggc	cgacgaacac	ctaattatca	4440
gagcaaaaat	ccagggtgagt	atactgtagg	gaattccatg	ttttacgatg	gtcctcaggt	4500
agcgaatctc	cagaacgtcg	acactggttt	ttggctggac	atgagcaatc	tctcagacgt	4560
tgtattatcc	agagagattc	aaacaggact	tcgagcacga	gctactttgg	aagaatccat	4620
gccgatgtta	gagaatttag	aagagcggtt	tagacgtttg	caagaaactt	gtgatgcggc	4680
tcgtactgag	atagaagaat	cgggatggac	tcgagagtc	gcatcaagaa	tggaaggcga	4740
tgaggcgcaa	ggaccttcta	gagcacaaca	agcttttcag	agctttgtaa	atgaatgtaa	4800
cagcatcgag	ttctcatttg	ggagctttgg	agagcatgtg	cgagttctct	gcgctagagt	4860
atcacgagga	ttagctcccg	caggagaggg	gattcgccgt	tgcttctctt	gttgtaaagg	4920
atcgacgcac	cgctacgctc	ctcgcatgga	cctatctcct	gaagggtgat	cgtagcaga	4980
gactttggct	agattcgcag	atgatattgg	aatagagcga	ggtgctgatg	gaacctacga	5040
tattcctttg	gtagatgatt	ggagaagagg	ggttcctagt	attgaaggag	aaggatctga	5100
ctcgatctat	gaaatcatga	tgcttatcta	tgaagttatg	gatattggatc	tagaaacacg	5160
aagatctttt	gcggtacagc	aagggcacta	tcaggaccca	agagcttcag	attatgacct	5220
cccacgtgct	agcgactatg	atttgcctag	aagcccatat	cctactccac	ctttgcctcc	5280
tagatatcag	ctacagaata	tggtatgtaga	agcagggttc	cgtagggcag	tttatgcttc	5340
ttttgtagca	ggaatgtaca	attatgtagt	gacacagccg	caagagcgta	ttcccaatag	5400
tcagcaggtg	gaagggtatc	tgctgtatat	gcttaccac	gggtcacaga	catttagaga	5460
cctgatgagg	cgttggaata	gagaagtcga	tagggaataa	actggtatct	accataggtt	5520

tgtagcaaaa aactaagccc accaagaaga aattctcttt ggtgggcttc tttttttatt  
caaaaaagaa agccctcttc a

5580  
5601

//

THIS PAGE BLANK (USPTO)

=> fil hcaplus  
 FILE 'HCAPLUS' ENTERED AT 08:37:27 ON 31 OCT 2001  
 USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
 PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
 COPYRIGHT (C) 2001 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications.

FILE COVERS 1947 - 31 Oct 2001 VOL 135 ISS 19  
 FILE LAST UPDATED: 30 Oct 2001 (20011030/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

This file supports REGISTRY for direct browsing and searching of all substance data from the REGISTRY file. Enter HELP FIRST for more information.

HCAplus now provides online access to patents and literature covered in CA from 1947 to the present. On April 22, 2001, bibliographic information and abstracts were added for over 2.2 million references published in CA from 1947 to 1966.

=>

=>

=> d stat que 15

```
L1      55 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("PROBST P"/AU OR "PROBST
      P"/IN OR "PROBST P A"/AU OR "PROBST P J"/AU) OR ("PROBST
      PETER"/AU OR "PROBST PETER"/IN OR "PROBST PETER G"/AU OR
      "PROBST PETER J"/AU OR "PROBST PETER S"/AU)
L2      356 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("BHATIA A"/AU OR "BHATIA A
      B"/AU OR "BHATIA A K"/AU OR "BHATIA A L"/AU OR "BHATIA A S"/AU
      OR "BHATIA AJAY"/AU OR "BHATIA AJAY"/IN)
L3      70 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("SKEIKY Y A W"/AU OR "SKEIKY
      YASIR"/AU OR "SKEIKY YASIR"/IN OR "SKEIKY YASIR A"/AU OR
      "SKEIKY YASIR A"/IN OR "SKEIKY YASIR A W"/AU OR "SKEIKY YASIR
      A W"/IN)
L4      22 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("FLING STEVEN"/AU OR "FLING
      STEVEN P"/AU OR "FLING STEVEN P"/IN OR "FLING STEVEN PAUL"/AU)
L5      2 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 AND L2 AND L3 AND L4
```

=>

=>

=> d ibib abs hitrn 15 1-2

```
L5  ANSWER 1 OF 2  HCAPLUS  COPYRIGHT 2001 ACS
ACCESSION NUMBER:  2001:417155  HCAPLUS
DOCUMENT NUMBER:   135:45174
TITLE:             Antigenic compounds and methods for treatment and
                   diagnosis of Chlamydial infection
```

**THIS PAGE BLANK (USPTO)**

INVENTOR(S): **Probst, Peter;** Bhatia, Ajay; Skeiky, Yasir  
 A. W.; Fling, Steven P.; Scholler, John  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 293 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001040474	A2	20010607	WO 2000-US32919	20001204
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
PRIORITY APPLN. INFO.:			US 1999-454684	A 19991203
			US 2000-556877	A 20000419
			US 2000-598419	A 20000620
AB Compds. and methods for the diagnosis and treatment of Chlamydial infection are disclosed. The compds. provided include polypeptides that contain at least one antigenic portion of a Chlamydia antigen and DNA sequences encoding such polypeptides from Chlamydia trachomatis and C. pneumoniae isolated using retroviral expression vector systems and subsequent immunol. anal. and epitope mapping. Pharmaceutical compns. and vaccines comprising such polypeptides or DNA sequences are also provided, together with antibodies directed against such polypeptides. Diagnostic kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of Chlamydial infection in patients and in biol. samples. In particular, fusion proteins are constructed from the Chlamydial proteins PmpA, PmpF, PmpH, PmpB, and PmpC fused with amino acid residues 192-323 of the Ra2 MTB32A serine proteinase from Mycobacterium tuberculosis.				

L5 ANSWER 2 OF 2 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 2000:402007 HCAPLUS

DOCUMENT NUMBER: 133:53686

TITLE: Chlamydial antigens and genomic DNA sequences for  
 treatment and diagnosis of chlamydial infection

INVENTOR(S): **Probst, Peter;** Bhatia, Ajay; Skeiky, Yasir  
 A. W.; Fling, Steven P.; Jen, Shyian; Stromberg, Erica  
 Jean

PATENT ASSIGNEE(S): Corixa Corporation, USA

SOURCE: PCT Int. Appl., 256 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000034483	A2	20000615	WO 1999-US29012	19991208

**THIS PAGE BLANK (USPTO)**

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,  
 CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,  
 IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
 MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,  
 SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,  
 AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,  
 DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,  
 CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

US 6166177 A 20001226 US 1998-208277 19981208

EP 1144642 A2 20011017 EP 1999-963037 19991208

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,  
 IE, SI, LT, LV, FI, RO

NO 2001002812 A 20010802 NO 2001-2812 20010607

PRIORITY APPLN. INFO.:

US 1998-208277 A 19981208

US 1999-288594 A 19990408

US 1999-410568 A 19991001

US 1999-426571 A 19991022

WO 1999-US29012 W 19991208

AB Compds. and methods for the diagnosis and treatment of Chlamydial infection are disclosed. The compds. provided include polypeptides that contain at least one antigenic portion of a Chlamydia antigen and DNA sequences encoding such polypeptides. Chlamydia antigens were isolated by expression cloning of a genomic DNA library of C. trachomatis LGV II, and shown to induce T cell proliferation and interferon-.beta. prodn. Immune responses of human PBMC and T cell lines are generated against the Chlamydia antigens. Pharmaceutical compns. and vaccines comprising such polypeptides or DNA sequences are also provided, together with antibodies directed against such polypeptides. Diagnostic kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of Chlamydial infection in patients and in biol. samples.

=>

=>

=> d stat que 16

L1 55 SEA FILE=HCAPLUS ABB=ON PLU=ON ("PROBST P"/AU OR "PROBST P"/IN OR "PROBST P A"/AU OR "PROBST P J"/AU) OR ("PROBST PETER"/AU OR "PROBST PETER"/IN OR "PROBST PETER G"/AU OR "PROBST PETER J"/AU OR "PROBST PETER S"/AU)

L2 356 SEA FILE=HCAPLUS ABB=ON PLU=ON ("BHATIA A"/AU OR "BHATIA A B"/AU OR "BHATIA A K"/AU OR "BHATIA A L"/AU OR "BHATIA A S"/AU OR "BHATIA AJAY"/AU OR "BHATIA AJAY"/IN)

L3 70 SEA FILE=HCAPLUS ABB=ON PLU=ON ("SKEIKY Y A W"/AU OR "SKEIKY YASIR"/AU OR "SKEIKY YASIR"/IN OR "SKEIKY YASIR A"/AU OR "SKEIKY YASIR A"/IN OR "SKEIKY YASIR A W"/AU OR "SKEIKY YASIR A W"/IN)

L4 22 SEA FILE=HCAPLUS ABB=ON PLU=ON ("FLING STEVEN"/AU OR "FLING STEVEN P"/AU OR "FLING STEVEN P"/IN OR "FLING STEVEN PAUL"/AU)

L5 2 SEA FILE=HCAPLUS ABB=ON PLU=ON L1 AND L2 AND L3 AND L4

L6 3 SEA FILE=HCAPLUS ABB=ON PLU=ON (L1 AND (L2 OR L3 OR L4)) NOT L5

=>

=>

**THIS PAGE BLANK (USPTO)**

=&gt; d ibib abs hitrn 16 1-3

L6 ANSWER 1 OF 3 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2001:780955 HCAPLUS  
 TITLE: Leishmania antigens for use in the therapy and  
 diagnosis of leishmaniasis  
 INVENTOR(S): Reed, Steven G.; Campos-Neto, Antonio; Webb, John R.;  
 Dillon, Davin C.; **Skeiky, Yasir A. W.**;  
**Bhatia, Ajay**; Coler, Rhea M.; **Probst,**  
**Peter**  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 193 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001079276	A2	20011025	WO 2001-US11254	20010405
W:	AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			
PRIORITY APPLN. INFO.:			US 2000-551974	A 20000414
			US 2000-565501	A 20000505
			US 2000-639206	A 20000814

AB Compositions and methods for preventing, treating and detecting leishmaniasis and stimulating immune responses in patients are disclosed. The compounds provided include polypeptides that contain at least an immunogenic portion of one or more Leishmania antigens, or a variant thereof. Vaccines and pharmaceutical compositions comprising such polypeptides, or polynucleotides encoding such polypeptides, are also provided and may be used, for example, for the prevention and therapy of leishmaniasis, as well as for the detection of Leishmania infection.

L6 ANSWER 2 OF 3 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2001:120546 HCAPLUS  
 DOCUMENT NUMBER: 134:325125  
 TITLE: CD8+ T cells recognize an inclusion  
 membrane-associated protein from the vacuolar pathogen  
 Chlamydia trachomatis  
 AUTHOR(S): **Fling, Steven P.**; Sutherland, R. Alec;  
 Steele, Lisa N.; Hess, Bruce; D'Orazio, Sarah E. F.;  
 Maisonneuve, Jean-Francois; Lampe, Mary F.;  
**Probst, Peter**; Starnbach, Michael N.  
 CORPORATE SOURCE: Corixa Corporation, Seattle, WA, 98104, USA  
 SOURCE: Proc. Natl. Acad. Sci. U. S. A. (2001), 98(3),  
 1160-1165  
 CODEN: PNASA6; ISSN: 0027-8424  
 PUBLISHER: National Academy of Sciences  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English

**THIS PAGE BLANK (USPTO)**

AB During infection with *Chlamydia trachomatis*, CD8+ T cells are primed, even though the bacteria remain confined to a host cell vacuole throughout their developmental cycle. Because CD8+ T cells recognize antigens processed from cytosolic proteins, the *Chlamydia* antigens recognized by these CD8+ T cells very likely have access to the host cell cytoplasm during infection. The identity of these *C. trachomatis* proteins has remained elusive, even though their localization suggests they may play important roles in the biol. of the organism. Here we use a retroviral expression system to identify Cap1, a 31-kDa protein from *C. trachomatis* recognized by protective CD8+ T cells. Cap1 contains no strong homol. to any known protein. Immunofluorescence microscopy by using Cap1-specific antibody demonstrates that this protein is localized to the vacuolar membrane. Cap1 is virtually identical among the human *C. trachomatis* serovars, suggesting that a vaccine incorporating Cap1 might enable the vaccine to protect against all *C. trachomatis* serovars. The identification of proteins such as Cap1 that assoc. with the inclusion membrane will be required to fully understand the interaction of *C. trachomatis* with its host cell.

REFERENCE COUNT: 34

REFERENCE(S): (1) Bannantine, J; Mol Microbiol 1998, V28, P1017 HCAPLUS  
 (2) Blattner, J; J Cell Biol 1992, V119, P1129 HCAPLUS  
 (4) Chakrabarti, S; Mol Cell Biol 1985, V5, P3403 HCAPLUS  
 (5) Denamur, E; J Gen Microbiol 1991, V137, P2525 HCAPLUS  
 (6) Earl, P; J Virol 1991, V65, P31 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 3 OF 3 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1997:692504 HCAPLUS

DOCUMENT NUMBER: 127:306395

TITLE: A Leishmania protein that modulates interleukin (IL)-12, IL-10, and tumor necrosis factor-.alpha. production and expression of B7-1 in human monocyte-derived antigen-presenting cells

AUTHOR(S): Probst, Peter; Skeiky, Yasir A. W.  
 ; Steeves, Michelle; Gervassi, Ana; Grabstein, Kenneth H.; Reed, Steven G.

CORPORATE SOURCE: Infectious Disease Research Inst., Seattle, WA, USA

SOURCE: Eur. J. Immunol. (1997), 27(10), 2634-2642

CODEN: EJIMAF; ISSN: 0014-2980

PUBLISHER: Wiley-VCH

DOCUMENT TYPE: Journal

LANGUAGE: English

AB LeIF, a gene homolog of the eukaryotic initiation factor 4A was 1st described as a leishmanial antigen that induced a Th1-type T cell response in peripheral blood mononuclear cells (PBMC) from leishmaniasis patients. The interferon (IFN)-.gamma. prodn. by PBMC was interleukin (IL)-12 dependent. The effects of LeIF were characterized on cytokine prodn. and expression of surface mols. by normal human monocytes as well as by monocyte-derived macrophages and dendritic cells (MoDC); LeIF was a strong inducer of IL-12 and, to a lesser extent, of IL-10, and tumor necrosis factor (TNF)-.alpha. in macrophages and MoDC. IL-12 prodn. did not require CD40 triggering, confirming that the ability of LeIF to induce IL-12 was, not mediated through an effect on T cells. However, addn. of sol. CD40 ligand (L) synergistically augmented IL-12 prodn. in macrophages and MoDC. The cytokine-inducing activity of LeIF is located in the N-terminal portion of the mol. and was both proteinase K sensitive and

**THIS PAGE BLANK (USPTO)**

polymyxin B resistant. LeIF, lipopolysaccharide, and fixed *Staphylococcus aureus* all induced comparable amts. of IL-12, validating the potent cytokine-inducing effects of LeIF. Of these stimuli, LeIF had the highest IL-12/IL-10 and IL-12/TNF- $\alpha$  ratio demonstrating the preference of LeIF for IL-12 induction. Studies investigating the expression of surface mols. showed that LeIF up-regulated B7-1 and CD54 (ICAM-1) on macrophages and MoDC. This is the 1st report describing IL-12 prodn., up-regulation of co-stimulatory and intercellular adhesion mols. by monocytic antigen-presenting cells in response to a protein from a pathogenic microorganism. These immunomodulatory characteristics of LeIF might be excellent properties for a Th11-type adjuvant.

=&gt;

=&gt;

=&gt; d stat que 17

```

L1      55 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("PROBST P"/AU OR "PROBST
        P"/IN OR "PROBST P A"/AU OR "PROBST P J"/AU) OR ("PROBST
        PETER"/AU OR "PROBST PETER"/IN OR "PROBST PETER G"/AU OR
        "PROBST PETER J"/AU OR "PROBST PETER S"/AU)
L2      356 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("BHATIA A"/AU OR "BHATIA A
        B"/AU OR "BHATIA A K"/AU OR "BHATIA A L"/AU OR "BHATIA A S"/AU
        OR "BHATIA AJAY"/AU OR "BHATIA AJAY"/IN)
L3      70 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("SKEIKY Y A W"/AU OR "SKEIKY
        YASIR"/AU OR "SKEIKY YASIR"/IN OR "SKEIKY YASIR A"/AU OR
        "SKEIKY YASIR A"/IN OR "SKEIKY YASIR A W"/AU OR "SKEIKY YASIR
        A W"/IN)
L4      22 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("FLING STEVEN"/AU OR "FLING
        STEVEN P"/AU OR "FLING STEVEN P"/IN OR "FLING STEVEN PAUL"/AU)
L5      2 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 AND L2 AND L3 AND L4
L6      3 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L1 AND (L2 OR L3 OR L4)) NOT
        L5
L7      4 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L2 AND (L3 OR L4)) NOT (L5
        OR L6)

```

=&gt;

=&gt;

=&gt; d ibib abs hitrn 17 1-4

L7 ANSWER 1 OF 4 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1999:534897 HCAPLUS

DOCUMENT NUMBER: 131:334383

TITLE: Cloning, characterization and serological evaluation  
of K9 and K26: two related hydrophilic antigens of  
*Leishmania chagasi*

AUTHOR(S): **Bhatia, A.**; Daifalla, N. S.; Jen, S.;  
Badaro, R.; Reed, S. G.; **Skeiky, Y. A. W.**

CORPORATE SOURCE: Corixa Corporation, Seattle, WA, USA

SOURCE: Mol. Biochem. Parasitol. (1999), 102(2), 249-261  
CODEN: MBIPDP; ISSN: 0166-6851

PUBLISHER: Elsevier Science Ireland Ltd.

DOCUMENT TYPE: Journal

LANGUAGE: English

AB We report here the mol. cloning and characterization of two related  
hydrophilic antigens of *Leishmania chagasi*. These two antigens have  
predicted mol. wts. of .apprx.9 and 26 kDa and detect antibodies in sera  
of patients with kala-azar (k). Thus, to maintain consistency with

**THIS PAGE BLANK (USPTO)**

nomenclature of the previously described 39kDa diagnostic antigen of *L. chagasi* (k39 [1]), these antigens are being referred to as k9 and k26. A significant difference between k9 and k26 is the presence of 11 copies of a 14 amino acid repeat in the open reading frame of k26. The region flanking the repeats of k26 shares a 69% identity with the open reading frame of k9. The recombinant proteins encoded by both antigens are very hydrophilic and show aberrant migration on SDS PAGE. Results of Southern blot anal. reveal that k9 and k26 are conserved to varying degrees among various *Leishmania* species. Interestingly, the repeat region of k26 is specific to *L. chagasi* and *L. donovani* while the flanking region is conserved among several other species. Transcript levels of k26 are significantly upregulated in the amastigote stage of the parasite. Our results show that recombinant K26 is specific in detecting antibodies in infection sera from visceral leishmaniasis (VL) patients. Thus rK26 may complement rK39 in a more accurate diagnosis of VL in the old and the new world.

REFERENCE COUNT: 29  
 REFERENCE(S): (1) Aly, R; Nucl Acids Res 1994, V22, P2922 HCAPLUS  
 (3) Badaro, R; J Infect Dis 1996, V173, P758 HCAPLUS  
 (4) Burns, J; Proc Natl Acad Sci USA 1993, V90, P775 HCAPLUS  
 (5) Chomczynski, P; Anal Biochem 1987, V162, P156 HCAPLUS  
 (6) Feinberg, A; Anal Biochem 1983, V132, P6 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 2 OF 4 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1997:415380 HCAPLUS

DOCUMENT NUMBER: 127:91226

TITLE: Molecular cloning and characterization of two iron superoxide dismutase cDNAs from *Trypanosoma cruzi*  
 AUTHOR(S): Ismail, Said O.; Paramchuk, Wendy; **Skeiky, Yasir A. W.**; Reed, Steven G.; **Bhatia, Ajay**;  
 Gedamu, Lashitew

CORPORATE SOURCE: Dep. Biological Sciences, Univ. Calgary, Calgary, AB, T2N 1N4, Can.

SOURCE: Mol. Biochem. Parasitol. (1997), 86(2), 187-197  
 CODEN: MBIPDP; ISSN: 0166-6851

PUBLISHER: Elsevier

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Two cDNAs (FeSODA and FeSODB cDNAs) corresponding to superoxide dismutase (1.15.1.1., SOD) were isolated from a *Trypanosoma cruzi* cDNA library. Comparison of the deduced amino acid sequences with previously reported SOD protein sequences revealed that the *T. cruzi* open reading frames had considerable homol. with FeSODs. The coding region of the *T. cruzi* FeSODB cDNA has been expressed in fusion with glutathione-S-transferase using an *Escherichia coli* mutant QC779, lacking both MnSOD and FeSOD genes (sodA sodB). Staining of native polyacrylamide gels for SOD activity of *T. cruzi* crude lysate and the recombinant SOD suggests that this protein is an FeSOD. The recombinant enzyme also protected the *E. coli* mutant QC779 from paraquat toxicity. Northern blot anal. showed that FeSODB is differentially expressed, showing a higher level at the epimastigote stage of *T. cruzi* development; whereas, FeSODA is constitutively expressed at a lower level in all developmental stages. Furthermore, Southern hybridization shows that both FeSODA and FeSODB genes appear to be present in the *T. cruzi* genome as multiple repeating units (multi-copy gene family).

**THIS PAGE BLANK (USPTO)**

L7 ANSWER 3 OF 4 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1995:781475 HCAPLUS

DOCUMENT NUMBER: 123:277157

TITLE: Molecular cloning, characterization, and expression in *Escherichia coli* of iron superoxide dismutase cDNA from *Leishmania donovani* chagasi. [Retraction to document cited in CA120:184186]AUTHOR(S): Ismail, Said O.; **Skeiky, Yasir A. W.**; **Bhatia, Ajay**; Omara-Opyene, Levi A.; Gedamu, Lashitew

CORPORATE SOURCE: Dep. Biol. Sci., Univ. Calgary, Calgary, AB, T2N 1N4, Can.

SOURCE: Infect. Immun. (1995), 63(9), 3749

CODEN: INFIBR; ISSN: 0019-9567

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The authors have requested that the article be retracted, as a *Trypanosoma cruzi* dDNA library rather than a *Leishmania donovani* chagasi dDNA library was used in screening.

L7 ANSWER 4 OF 4 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1994:184186 HCAPLUS

DOCUMENT NUMBER: 120:184186

TITLE: Molecular cloning, characterization, and expression in *Escherichia coli* of iron superoxide dismutase cDNA from *Leishmania donovani* chagasiAUTHOR(S): Ismail, Said O.; **Skeiky, Yasir A. W.**; **Bhatia, Ajay**; Omara-Opyene, Levi A.; Gedamu, Lashitew

CORPORATE SOURCE: Dep. Biol. Sci., Univ. Calgary, Calgary, AB, T2N 1N4, Can.

SOURCE: Infect. Immun. (1994), 62(2), 657-64

CODEN: INFIBR; ISSN: 0019-9567

DOCUMENT TYPE: Journal

LANGUAGE: English

AB A cDNA corresponding to superoxide dismutase (SOD; EC 1.15.1.1.) was isolated from a *Leishmania donovani* chagasi (L. d. chagasi) promastigote cDNA library, using PCR with a set of primers derived from conserved amino acids of manganese SODs (MnSODs) and iron SODs (FeSODs). Comparison of the deduced amino acid sequences with previously reported SOD amino acid sequences revealed that the L. d. chagasi 585-bp open reading frame had considerable homol. with FeSODs and MnSODs. The highest homol. was shared with prokaryotic FeSODs. The coding region of L. d. chagasi SOD cDNA has been expressed in fusion with glutathione-S-transferase, using an *Escherichia coli* mutant, QC779, lacking both MnSOD and FeSOD genes (sodA and sodB). Staining of native polyacrylamide gels for SOD activity of *Leishmania* crude lysate and the recombinant SOD revealed that both had SOD activity that was inactivated by 5 mM hydrogen peroxide but not by 2 mM potassium cyanide, which is indicative of FeSOD. The recombinant enzyme also protected *E. coli* mutant QC779 from paraquat toxicity. This indicated that the glutathione-S-transferase peptide does not interfere with the in vivo and in vitro activities of the recombinant SOD. Cross-species hybridization showed that FeSOD is highly conserved in the *Leishmania* genus. Interestingly, the hybridization pattern of the FeSOD gene(s) coincided with other classification schemes that divide *Leishmania* species into complexes. The cloning of FeSOD cDNA may contribute to the understanding of the role of SODs in *Leishmania* pathogenesis.

**THIS PAGE BLANK (USPTO)**

=>  
=>

=> d stat que 18

```
L1      55 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("PROBST P"/AU OR "PROBST
P"/IN OR "PROBST P A"/AU OR "PROBST P J"/AU) OR ("PROBST
PETER"/AU OR "PROBST PETER"/IN OR "PROBST PETER G"/AU OR
"PROBST PETER J"/AU OR "PROBST PETER S"/AU)
L2      356 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("BHATIA A"/AU OR "BHATIA A
B"/AU OR "BHATIA A K"/AU OR "BHATIA A L"/AU OR "BHATIA A S"/AU
OR "BHATIA AJAY"/AU OR "BHATIA AJAY"/IN)
L3      70 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("SKEIKY Y A W"/AU OR "SKEIKY
YASIR"/AU OR "SKEIKY YASIR"/IN OR "SKEIKY YASIR A"/AU OR
"SKEIKY YASIR A"/IN OR "SKEIKY YASIR A W"/AU OR "SKEIKY YASIR
A W"/IN)
L4      22 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("FLING STEVEN"/AU OR "FLING
STEVEN P"/AU OR "FLING STEVEN P"/IN OR "FLING STEVEN PAUL"/AU)
L5      2 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 AND L2 AND L3 AND L4
L6      3 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L1 AND (L2 OR L3 OR L4)) NOT
L5
L8      0 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L3 AND L4) NOT (L5 OR L6)
```

=> d stat que 112

```
L1      55 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("PROBST P"/AU OR "PROBST
P"/IN OR "PROBST P A"/AU OR "PROBST P J"/AU) OR ("PROBST
PETER"/AU OR "PROBST PETER"/IN OR "PROBST PETER G"/AU OR
"PROBST PETER J"/AU OR "PROBST PETER S"/AU)
L2      356 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("BHATIA A"/AU OR "BHATIA A
B"/AU OR "BHATIA A K"/AU OR "BHATIA A L"/AU OR "BHATIA A S"/AU
OR "BHATIA AJAY"/AU OR "BHATIA AJAY"/IN)
L3      70 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("SKEIKY Y A W"/AU OR "SKEIKY
YASIR"/AU OR "SKEIKY YASIR"/IN OR "SKEIKY YASIR A"/AU OR
"SKEIKY YASIR A"/IN OR "SKEIKY YASIR A W"/AU OR "SKEIKY YASIR
A W"/IN)
L4      22 SEA FILE=HCAPLUS ABB=ON  PLU=ON  ("FLING STEVEN"/AU OR "FLING
STEVEN P"/AU OR "FLING STEVEN P"/IN OR "FLING STEVEN PAUL"/AU)
L5      2 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 AND L2 AND L3 AND L4
L6      3 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L1 AND (L2 OR L3 OR L4)) NOT
L5
L7      4 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L2 AND (L3 OR L4)) NOT (L5
OR L6)
L11     30 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L1 OR L2 OR L3 OR L4 OR L5)
AND ((INFECT? OR ?CHLAMYD? OR BACTER? OR VIRUS?) (L) (DIAGNOSIS
OR DX OR DIAGNOS?))
L12     27 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L11 NOT (L5 OR L6 OR L7)
```

=>  
=>

=> d ibib abs hitrn 112 1-27

```
L12 ANSWER 1 OF 27 HCAPLUS COPYRIGHT 2001 ACS
ACCESSION NUMBER: 2001:687327 HCAPLUS
DOCUMENT NUMBER: 135:271872
TITLE: Compounds and methods for immunotherapy and diagnosis
of tuberculosis
INVENTOR(S): Reed, Steven G.; Skeiky, Yasir A. W.;
```

**THIS PAGE BLANK (USPTO)**

PATENT ASSIGNEE(S): Dillon, Davin C.; Campos-Neto, Antonio; Houghton, Raymond; Vedvick, Thomas S.; Twardzik, Daniel R.  
 SOURCE: Corixa Corporation, USA  
 U.S., 100 pp., Cont.-in-part of U.S. Ser. No. 730,510.  
 CODEN: USXXAM  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 3  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 6290969	B1	20010918	US 1997-818112	19970313
ZA 9607394	A	19970505	ZA 1996-7394	19960830
WO 9816646	A2	19980423	WO 1997-US18293	19971007
WO 9816646	A3	19981008		
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
AU 9748144	A1	19980511	AU 1997-48144	19971007
EP 932681	A2	19990804	EP 1997-910873	19971007
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
BR 9712518	A	20001024	BR 1997-12518	19971007
JP 2001501832	T2	20010213	JP 1998-518456	19971007
NO 9901694	A	19990610	NO 1999-1694	19990409
PRIORITY APPLN. INFO.:				
US 1995-523436 B2 19950901				
US 1995-533634 B2 19950922				
US 1996-620874 B2 19960322				
US 1996-659683 A2 19960605				
US 1996-680574 A2 19960712				
US 1996-730510 A2 19961011				
US 1997-818112 A 19970313				
WO 1997-US18293 W 19971007				
AB Compds. and methods for inducing protective immunity against tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one immunogenic portion of one or more Mycobacterium tuberculosis proteins and DNA mols. encoding such polypeptides. Such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against M. tuberculosis <b>infection</b> , or may be used for the <b>diagnosis</b> of tuberculosis.				
REFERENCE COUNT: 61				
REFERENCE(S):				
(2) Andersen; Infection and Immunity 1993, V61(3), P844 HCAPLUS				
(3) Andersen; Scand J Immunol 1992, V36, P823 HCAPLUS				
(4) Andersen, P; Infection and Immunity 1994, V62(6), P2536 HCAPLUS				
(5) Anon; WO 8805823 1988 HCAPLUS				
(6) Anon; EP 419355 A1 1991 HCAPLUS				
ALL CITATIONS AVAILABLE IN THE RE FORMAT				
L12 ANSWER 2 OF 27 HCAPLUS COPYRIGHT 2001 ACS				
ACCESSION NUMBER: 2001:636186 HCAPLUS				
DOCUMENT NUMBER: 135:209890				

**THIS PAGE BLANK (USPTO)**

TITLE: Compounds and methods for diagnosis and immunotherapy of tuberculosis  
 INVENTOR(S): Campos-Neto, Antonio; **Skeiky, Yasir**; Ovendale, Pamela; Jen, Shyian; Lodes, Michael  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 161 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001062893	A2	20010830	WO 2001-US5992	20010226
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
PRIORITY APPLN. INFO.:			US 2000-185037	P 20000225
			US 2000-223828	P 20000808

AB Compsds. and methods for **diagnosing** tuberculosis or for inducing protective immunity against tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one immunogenic portion of one or more Mycobacterium proteins and DNA mols. encoding such polypeptides. **Diagnostic** kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of Mycobacterium **infection** in patients and biol. samples. Antibodies directed against such polypeptides are also provided. In addn., such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against Mycobacterium **infection**.

L12 ANSWER 3 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2001:582322 HCAPLUS  
 DOCUMENT NUMBER: 135:179697  
 TITLE: Diagnosis of tuberculosis  
 INVENTOR(S): Alderson, Mark R.; Dillon, Davin C.; **Skeiky, Yasir A. w.**; Campos-Neto, Antonio  
 PATENT ASSIGNEE(S): USA  
 SOURCE: U.S. Pat. Appl. Publ., 74 pp., Cont.-in-part of U.S. Ser. No. 858,998, abandoned.  
 CODEN: USXXCO  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 2  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2001012888	A1	20010809	US 1998-73009	19980505
WO 9853076	A2	19981126	WO 1998-US10514	19980520
WO 9853076	A3	19990401		
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP,				

**THIS PAGE BLANK (USPTO)**

KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO,  
 NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA,  
 UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM  
 RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES,  
 FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI,  
 CM, GA, GN, ML, MR, NE, SN, TD, TG

AU 9875916	A1	19981211	AU 1998-75916	19980520
BR 9809443	A	20000613	BR 1998-9443	19980520
NO 9905690	A	20000118	NO 1999-5690	19991119
PRIORITY APPLN. INFO.:			US 1997-858998	B2 19970520
			US 1998-73009	A 19980505
			WO 1998-US10514	W 19980520

AB Compds. and methods for **diagnosing** tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one antigenic portion of one or more Mycobacterium tuberculosis proteins, and DNA sequences encoding such polypeptides. **Diagnostic** kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of M. tuberculosis **infection** in patients and biol. samples. Antibodies directed against such polypeptides are also provided.

L12 ANSWER 4 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 2001:550008 HCAPLUS  
 TITLE: Serological expression cloning and immunological evaluation of MTB48, a novel mycobacterium tuberculosis antigen  
 AUTHOR(S): Lodes, Michael J.; Dillon, Davin C.; Mohamath, Raodoh; Day, Craig H.; Benson, Darin R.; Reynolds, Lisa D.; McNeill, Patricia; Sampaio, Diana Pedral; **Skeiky, Yasir A. W.**; Badaro, Roberto; Persing, David H.; Reed, Steven G.; Houghton, Raymond L.  
 CORPORATE SOURCE: Corixa Corporation, Seattle, WA, 98104, USA  
 SOURCE: J. Clin. Microbiol. (2001), 39(7), 2485-2493  
 CODEN: JCMIDW; ISSN: 0095-1137  
 PUBLISHER: American Society for Microbiology  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English

AB Improved **diagnostics** are needed for the detection of Mycobacterium tuberculosis, esp. for patients with smear-neg. disease. To address this problem, we have screened M. tuberculosis (H37Rv and Erdman strains) genomic expression libraries with pooled sera from patients with extrapulmonary disease and with sera from patients with elevated reactivity with M. tuberculosis lysate. Both serum pools were reactive with clones expressing a recombinant protein referred to here as MTB48. The genomic sequence of the resulting clones was identical to that of the M. tuberculosis H37Rv isolate and showed 99% identity to the Mycobacterium bovis and M. bovis BCG isolate sequences. The genomic location of this sequence is 826 bp upstream of a region contg. the esat-6 gene that is deleted in the M. bovis BCG isolate. The mtb48 1,380-bp open reading frame encodes a predicted 47.6-kDa polypeptide with no known function. Southern and Western blot analyses indicate that this sequence is present in a single copy and is conserved in the M. tuberculosis and M. bovis isolates tested but not in other mycobacterial species tested, including Mycobacterium leprae and Mycobacterium avium. In addn., the native protein was detected in the cytoplasm, as was a processed form that was also shed into the medium during culture. Serol. anal. of recombinant MTB48 and the M. tuberculosis 38-kDa antigen with a panel of patient and control sera indicates that the inclusion of recombinant MTB48 in a prototype serodiagnostic test increases assay sensitivity for M.

**THIS PAGE BLANK (USPTO)**

tuberculosis **infection** when it is combined with other known immunodominant antigens, such as the 39-kDa antigen.

REFERENCE COUNT: 35  
 REFERENCE(S): (1) Andersen, A; Infect Immun 1989, V57, P2481 HCAPLUS  
 (2) Ashbridge, K; J Immunol 1990, V144, P3137 HCAPLUS  
 (9) Dillon, D; J Clin Microbiol 2000, V38, P3285 HCAPLUS  
 (18) Harboe, M; J Infect Dis 1992, V166, P874 HCAPLUS  
 (20) Hendrickson, R; J Clin Microbiol 2000, V38, P2354 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 5 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2001:331212 HCAPLUS  
 DOCUMENT NUMBER: 134:352261  
 TITLE: Compounds and methods for the detection and prevention of T. cruzi infection  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**; Lodes, Michael J.; Houghton, Raymond L.; Smith, John M.; McNeill, Patricia D.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: U.S., 56 pp., Cont.-in-part of U.S. 6,054,135.  
 CODEN: USXXAM  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 4  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 6228372	B1	20010508	US 1997-993674	19971218
US 6054135	A	20000425	US 1997-834306	19970415
WO 9931246	A1	19990624	WO 1998-US25871	19981204
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 9918057	A1	19990705	AU 1999-18057	19981204
EP 1038000	A1	20000927	EP 1998-962923	19981204
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				

PRIORITY APPLN. INFO.: US 1997-834306 A2 19970415  
 WO 1996-US18624 A2 19961114  
 US 1997-993674 A 19971218  
 WO 1998-US25871 W 19981204

AB Compds. and methods are provided for **diagnosing** Trypanosoma cruzi **infection**. The disclosed compds. are polypeptides, or antibodies thereto, that contain one or more epitopes of T. cruzi antigens. The compds. are useful in a variety of immunoassays for detecting T. cruzi **infection**. The polypeptide compds. are further useful in vaccines and pharmaceutical compns. for inducing protective immunity against Chagas' disease in individuals exposed to T. cruzi.

REFERENCE COUNT: 23  
 REFERENCE(S): (2) Anon; WO 9209895 1992 HCAPLUS

**THIS PAGE BLANK (USPTO)**

(3) Anon; WO 9316199 1993 HCAPLUS  
 (4) Anon; WO 9401776 1994 HCAPLUS  
 (5) Anon; WO 9629605 1996 HCAPLUS  
 (6) Anon; WO 9718475 1997 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 6 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2001:265269 HCAPLUS  
 DOCUMENT NUMBER: 134:309685  
 TITLE: Fusion proteins of Mycobacterium tuberculosis  
 INVENTOR(S): **Skeiky, Yasir**; Reed, Steven; Houghton,  
 Raymond L.; Mcneill, Patricia D.; Dillon, Davin C.;  
 Lodes, Michael L.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 168 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001024820	A1	20010412	WO 2000-US28095	20001010
W:	AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			

PRIORITY APPLN. INFO.: US 1999-158338 P 19991007  
 US 1999-158425 P 19991007

AB The present invention relates to fusion proteins contg. at least two Mycobacterium species antigens. In particular, it relates to nucleic acids encoding fusion proteins that include two or more individual M. tuberculosis antigens, which increase serol. sensitivity of sera from individuals **infected** with tuberculosis, and methods for their use in the **diagnosis**, treatment, and prevention of tuberculosis **infection**.

REFERENCE COUNT: 9  
 REFERENCE(S): (1) Alderson; J Exp Med 2000, V191(3), P551 HCAPLUS  
 (2) Brandt; Infect Immun 2000, V68(2), P791 HCAPLUS  
 (3) Coler; J Immunol 1998, V161(5), P2356 HCAPLUS  
 (4) Corixa Corporation; WO 9951748 A2 1999 HCAPLUS  
 (5) Hendrickson; J Clin Microbiol 2000, V38(6), P2354 HCAPLUS

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 7 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2000:698208 HCAPLUS  
 DOCUMENT NUMBER: 134:235797  
 TITLE: Molecular and immunological characterization of Mycobacterium tuberculosis CFP-10, an immunodiagnostic antigen missing in Mycobacterium bovis BCG  
 AUTHOR(S): Dillon, Davin C.; Alderson, Mark R.; Day, Craig H.; Bement, Teresa; Campos-Neto, Antonio; **Skeiky**,

**THIS PAGE BLANK (USPTO)**

Yasir A. W.; Vedvick, Thomas; Badaro, Roberto;  
 Reed, Steven G.; Houghton, Raymond  
 CORPORATE SOURCE: Corixa Corporation, Seattle, WA, 98104, USA  
 SOURCE: J. Clin. Microbiol. (2000), 38(9), 3285-3290  
 CODEN: JCMIDW; ISSN: 0095-1137  
 PUBLISHER: American Society for Microbiology  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English

AB In order to identify antigens that may be used in the serodiagnosis of active tuberculosis (TB), we screened a Mycobacterium tuberculosis genomic expression library with a pool of sera from patients **diagnosed** with active pulmonary TB. The sera used lacked reactivity with a recombinant form of the M. tuberculosis 38-kDa antigen (r38kDa), and the goal was to identify antigens that might complement r38kDa in a serodiagnostic assay. Utilizing this strategy, we identified a gene, previously designated lhp, which encodes a 100-amino-acid protein referred to as culture filtrate protein 10 (CFP-10). The lhp gene is located directly upstream of esat-6, within a region missing in M. bovis BCG. Immunoblot anal. demonstrated that CFP-10 is present in M. tuberculosis CFP, indicating that it is likely a secreted or shed antigen. Purified recombinant CFP-10 (rCFP-10) was shown to be capable of detecting specific antibody in a percentage of TB patients that lack reactivity with r38kDa, most notably in smear-neg. cases, where sensitivity was increased from 21% for r38kDa alone to 40% with the inclusion of rCFP-10. In smear-pos. patient sera, sensitivity was increased from 49% for r38kDa alone to 58% with the inclusion of rCFP-10. In addn., rCFP-10 was shown to be a potent T-cell antigen, eliciting proliferative responses and gamma interferon prodn. from peripheral blood mononuclear cells in 70% of purified protein deriv.-pos. individuals without evident disease. The responses to this antigen argue for the inclusion of rCFP-10 in a polyvalent serodiagnostic test for detection of active TB **infection**. RCFP-10 could also contribute to the development of a recombinant T-cell **diagnostic** test capable of detecting exposure to M. tuberculosis.

REFERENCE COUNT: 25

REFERENCE(S): (1) Alderson, M; J Exp Med 2000, V191, P551 HCAPLUS  
 (2) Amara, R; Infect Immun 1996, V64, P3765 HCAPLUS  
 (3) Andersen, A; Infect Immun 1989, V57, P2481 HCAPLUS  
 (4) Berthet, F; Microbiology 1998, V144, P3195 HCAPLUS  
 (6) Chang, Z; J Biol Chem 1994, V269, P1956 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 8 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 2000:609008 HCAPLUS

DOCUMENT NUMBER: 133:190202

TITLE: Compounds and methods for the detection and prevention of Trypanosoma cruzi infection

INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**; Lodes, Michael J.; Houghton, Raymond L.; Smith, John M.; McNeill, Patricia D.

PATENT ASSIGNEE(S): Corixa Corporation, USA

SOURCE: PCT Int. Appl., 55 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
-----	----	-----	-----	-----

**THIS PAGE BLANK (USPTO)**

WO 2000050897 A1 20000831 WO 2000-US4815 20000224

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,  
 CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,  
 IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
 MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,  
 SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,  
 AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,  
 DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,  
 CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

US 1999-256976 A 19990224

AB Compds. and methods are provided for **diagnosing** Trypanosoma cruzi **infection**. The disclosed compds. are polypeptides, or antibodies thereto, that contain one or more epitopes of T. cruzi antigens. The compds. are useful in a variety of immunoassays for detecting T. cruzi **infection**. The polypeptide compds. are further useful in vaccines and pharmaceutical compns. for inducing protective immunity against Chagas' disease in individuals exposed to T. cruzi.

REFERENCE COUNT:

7

REFERENCE(S):

- (3) Corixa Corp; WO 9629605 A 1996 HCAPLUS
  - (4) Corixa Corp; WO 9718475 A 1997 HCAPLUS
  - (5) Corixa Corp; WO 9931246 A 1999 HCAPLUS
  - (6) Houghton, R; THE JOURNAL OF INFECTIOUS DISEASES 2000, V181(1), P325 HCAPLUS
  - (7) Reed, S; WO 9316199 A 1993 HCAPLUS
- ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 9 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 2000:588348 HCAPLUS

DOCUMENT NUMBER: 134:290933

TITLE: Cloning of a Mycobacterium tuberculosis gene encoding a purified protein derivative protein that elicits strong tuberculosis-specific delayed-type hypersensitivity

AUTHOR(S): Coler, Rhea N.; **Skeiky, Yasir A. W.**;  
 Ovendale, Pamela J.; Vedvick, Thomas S.; Gervassi, Luis; Guderian, Jeff; Jen, Shyian; Reed, Steven G.; Campos-Neto, Antonio

CORPORATE SOURCE: Infectious Disease Research Institute, Seattle, WA, 98104, USA

SOURCE: J. Infect. Dis. (2000), 182(1), 224-233

CODEN: JIDIAQ; ISSN: 0022-1899

PUBLISHER: University of Chicago Press

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The purified protein deriv. (PPD) skin test has been used for the **diagnosis** of tuberculosis for more than 75 yr. However, the test lacks specificity because all mycobacteria share antigens present in PPD. Therefore, sensitization with nontuberculous pathogenic or with environmental nonpathogenic mycobacteria can lead to pos. skin tests. This communication describes a novel PPD protein present only in tuberculous complex mycobacteria. A recombinant protein was obtained and named DPPD on the basis of the first 4 amino acids of its N-terminus sequence. DPPD elicited delayed-type hypersensitivity (DTH) in 100% of Mycobacterium tuberculosis-**infected** guinea pigs but in no animals sensitized with several organisms representative of all members of the Mycobacterium genus. Preliminary results indicate that DPPD induces strong and specific DTH in humans. This work points to the definition of

**THIS PAGE BLANK (USPTO)**

a single recombinant M. tuberculosis protein that may be an alternative to the PPD test.

REFERENCE COUNT: 44  
 REFERENCE(S): (1) Affronti, L; J Biol Stand 1986, V14, P177 HCAPLUS  
 (2) Affronti, L; J Biol Stand 1986, V14, P187 HCAPLUS  
 (7) Coler, R; J Immunol 1998, V161, P2356 HCAPLUS  
 (9) Daniel, T; Microbiol Rev 1978, V42, P84 HCAPLUS  
 (11) Feinberg, A; Anal Biochem 1983, V132, P6 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 10 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 2000:271863 HCAPLUS  
 DOCUMENT NUMBER: 132:307235  
 TITLE: Combination polypeptide compounds for the detection of Trypanosoma cruzi infection  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**; Lodes, Michael J.; Houghton, Raymond L.  
 PATENT ASSIGNEE(S): Corixa, USA  
 SOURCE: U.S., 50 pp., Cont.-in-part of Appl. No. PCT/US96/18624.  
 CODEN: USXXAM  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 4  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 6054135	A	20000425	US 1997-834306	19970415
WO 9718475	A1	19970522	WO 1996-US18624	19961114
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: KE, LS, MW, SD, SZ, UG, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 6228372	B1	20010508	US 1997-993674	19971218
PRIORITY APPLN. INFO.:			WO 1996-US18624	A2 19961114
			US 1995-557309	A 19951114
			US 1997-834306	A2 19970415

AB Compds. are provided for **diagnosing** Trypanosoma cruzi **infection**. The disclosed compds. are combination polypeptides that contain a peptide corresponding to novel gene with a repeat sequence and one or more epitopes of T. cruzi antigens such as TcD, TcE, or PEP-2 epitope. The compds. are useful in a variety of immunoassays for detecting T. cruzi **infection**. Peptide epitopes of Trypanosoma cruzi have been identified through expression cloning. A tripeptide (2/D/E) contg. three epitopes (TcD, TcE, PEP-2) was used in ELISA to detect antibodies to T. cruzi in 239 of 240 consensus-pos. sera and 41 of 42 sera confirmed pos. by radioimmunopptn. assay. The 1 discrepant consensus-pos. serum was used to expression-clone a novel T. cruzi gene that contained a repeat sequence. A peptide corresponding to one reading frame of this sequence, TcLol.2, was reactive with T. cruzi-pos. sera and enhanced reactivity of other peptides. Thus, peptides and recombinants contg. multiple repeat epitopes are powerful tools for developing assays for T. cruzi antibody detection and have direct application in blood screening.

**THIS PAGE BLANK (USPTO)**

REFERENCE COUNT: 16  
 REFERENCE(S): (3) Anon; WO 9209895 1992 HCAPLUS  
 (4) Anon; WO 9316199 1993 HCAPLUS  
 (5) Anon; WO 9401776 1994 HCAPLUS  
 (6) Anon; WO 9629605 1996 HCAPLUS  
 (7) Campetella, O; Parasitology Today V8(11), P378  
 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 11 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1999:659510 HCAPLUS  
 DOCUMENT NUMBER: 131:296204  
 TITLE: Fusion proteins of Mycobacterium tuberculosis antigens  
 containing domains from more than one Mycobacterium  
 protein and their uses  
 INVENTOR(S): Skeiky, Yasir A. W.; Alderson, Mark;  
 Campos-Neto, Antonio  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 83 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9951748	A2	19991014	WO 1999-US7717	19990407
WO 9951748	A3	20000203		
W:	AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			
AU 9934817	A1	19991025	AU 1999-34817	19990407
EP 1068329	A2	20010117	EP 1999-916513	19990407
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI			
BR 9909472	A	20010911	BR 1999-9472	19990407
NO 2000005050	A	20001130	NO 2000-5050	20001006
PRIORITY APPLN. INFO.:			US 1998-56556	A 19980407
			US 1998-223040	A 19981230
			WO 1999-US7717	W 19990407

AB Fusion proteins contg. antigenic regions from two or more proteins (up to five) of Mycobacterium tuberculosis that can be used in the **diagnosis**, treatment and prevention of tuberculosis **infection** are described. These fusion proteins retain the antigenicity of the originals. A series of twelve fusion proteins contg. combinations of peptides from M. tuberculosis antigens were constructed by std. methods and manufd. as inclusion bodies in Escherichia coli. The fusion proteins stimulated T cell proliferation in PPD+ patients with proliferation patterns similar to those of the individual components. Immunization of mice with the fusion proteins induced strong interferon .gamma. and interleukin 4 responses with the strength of the responses depending upon the adjuvant used.

**THIS PAGE BLANK (USPTO)**

L12 ANSWER 12 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1999:549166 HCAPLUS  
 DOCUMENT NUMBER: 131:166234  
 TITLE: Antigenic proteins and their genes for use in  
 diagnosis of tuberculosis  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**;  
 Dillon, Davin C.; Campos-Neto, Antonio; Houghton,  
 Raymond; Vedvick, Thomas S.; Twardzik, Daniel R.;  
 Lodes, Michael J.; Hendrickson, Ronald C.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 323 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9942118	A2	19990826	WO 1999-US3265	19990217
W:	AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			
AU 9926819	A1	19990906	AU 1999-26819	19990217
EP 1091749	A1	20010418	EP 1999-907065	19990217
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI			
PRIORITY APPLN. INFO.:			US 1998-24753 A 19980218	
			US 1998-72596 A 19980505	
			WO 1999-US3265 W 19990217	

AB Numerous antigenic protein fragments are isolated from Mycobacterium tuberculosis culture filtrates and patient sera and used to isolate genomic fragments of the M. tuberculosis DNA. These substances provide reagents for the **diagnosis** of **infection** by detecting antibodies against M. tuberculosis, **diagnosis** of tuberculosis using PCR primers or hybridization probes, and serodiagnosis or immunoassay of tuberculosis. The construction and use of tuberculosis fusion proteins for serodiagnosis are also provided.

L12 ANSWER 13 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1999:549126 HCAPLUS  
 DOCUMENT NUMBER: 131:183862  
 TITLE: Compounds and methods for immunotherapy and diagnosis  
 of tuberculosis  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**;  
 Dillon, Davin C.; Campos-Neto, Antonio; Houghton,  
 Raymond; Vedvick, Thomas S.; Twardzik, Daniel R.;  
 Lodes, Michael J.; Hendrickson, Ronald C.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 299 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English

**THIS PAGE BLANK (USPTO)**

FAMILY ACC. NUM. COUNT: 1  
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9942076	A2	19990826	WO 1999-US3268	19990217
WO 9942076	A3	19991014		
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 9927663	A1	19990906	AU 1999-27663	19990217
EP 1071451	A2	20010131	EP 1999-908169	19990217
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				

PRIORITY APPLN. INFO.:  
US 1998-25197 A 19980218  
US 1998-72967 A 19980505  
WO 1999-US3268 W 19990217

AB Compsds. and methods for inducing protective immunity against tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one immunogenic portion of one or more Mycobacterium tuberculosis proteins and DNA mols. encoding such polypeptides. Such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against M. tuberculosis **infection**, or may be used for the **diagnosis** of tuberculosis.

L12 ANSWER 14 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
ACCESSION NUMBER: 1999:537911 HCAPLUS  
DOCUMENT NUMBER: 131:169281  
TITLE: Compounds and methods for the detection of Trypanosoma cruzi infection  
INVENTOR(S): Reed, Steven G.; Houghton, Raymond; **Skeiky, Yasir A. W.**  
PATENT ASSIGNEE(S): Corixa Corporation, USA  
SOURCE: U.S., 26 pp., Cont.-in-part of U. S. 5,756,662.  
CODEN: USXXAM  
DOCUMENT TYPE: Patent  
LANGUAGE: English  
FAMILY ACC. NUM. COUNT: 2  
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 5942403	A	19990824	US 1997-929414	19970915
US 5756662	A	19980526	US 1995-403379	19950314
CA 2215104	AA	19960926	CA 1996-2215104	19960312

PRIORITY APPLN. INFO.: US 1995-403379 19950314

AB Compsds. and methods for **diagnosing** Trypanosoma cruzi **infection**, or for screening for T. cruzi or Leishmania **infection**, are disclosed. The disclosed compds. are polypeptides, or antibodies thereto, that contain one or more antigenic epitopes of T. cruzi proteins. The compds. are useful in a variety of immunoassays for detecting T. cruzi **infection**. The polypeptide compds. are further useful in vaccines and pharmaceutical compns. for preventing Chagas' disease in individuals exposed to T. cruzi.

**THIS PAGE BLANK (USPTO)**

REFERENCE COUNT: 20  
 REFERENCE(S): (2) Anon; WO 92/09895 1992 HCAPLUS  
 (3) Anon; WO 93/16199 1993 HCAPLUS  
 (4) Anon; WO 94/01776 1994 HCAPLUS  
 (5) Burns; Proc Nat Acad Sci USA 1992, V89, P1239  
 HCAPLUS  
 (6) Buschiazzi; Molecular and Biochemical Parasitology  
 1992, V54, P125 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 15 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1999:405092 HCAPLUS  
 DOCUMENT NUMBER: 131:54723  
 TITLE: Peptides and antibodies for detection and prevention  
 of Trypanosoma cruzi infection  
 INVENTOR(S): Reed, Steven G.; Skeiky, Yasir A. W.; Lodes,  
 Michael J.; Houghton, Raymond L.; Smith, John M.;  
 McNeill, Patricia D.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 103 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 4  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9931246	A1	19990624	WO 1998-US25871	19981204
W:	AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			
US 6228372	B1	20010508	US 1997-993674	19971218
AU 9918057	A1	19990705	AU 1999-18057	19981204
EP 1038000	A1	20000927	EP 1998-962923	19981204
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI			

PRIORITY APPLN. INFO.: US 1997-993674 A 19971218  
 US 1997-834306 A2 19970415  
 WO 1998-US25871 W 19981204

AB Polypeptides, or antibodies thereto, that contain one or more epitopes of Trypanosoma cruzi antigens are provided for **diagnosing** T. cruzi **infection**. The compds. are useful in a variety of immunoassays for detecting T. cruzi **infection**. The polypeptide compds. are further useful in vaccines and pharmaceutical compns. for inducing protective immunity against Chagas' disease in individuals exposed to T. cruzi.

REFERENCE COUNT: 3  
 REFERENCE(S): (1) Corixa Corp; WO 9629605 A 1996 HCAPLUS  
 (2) Corixa Corp; WO 9718475 A 1997 HCAPLUS  
 (3) Reed, S; WO 9316199 A 1993 HCAPLUS

L12 ANSWER 16 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1999:350405 HCAPLUS

**THIS PAGE BLANK (USPTO)**

DOCUMENT NUMBER: 131:156679  
 TITLE: Molecular characterization and human T-cell responses to a member of a novel Mycobacterium tuberculosis mtb39 gene family  
 AUTHOR(S): Dillon, Davin C.; Alderson, Mark R.; Day, Craig H.; Lewinsohn, David M.; Coler, Rhea; Bement, Teresa; Campos-Neto, Antonio; **Skeiky, Y. A. W.**; Orme, Ian M.; Roberts, Alan; Steen, Sean; Dalemans, Wilfried; Badaro, Roberto; Reed, Steven G.  
 CORPORATE SOURCE: Corixa Corporation, Seattle, WA, 98104, USA  
 SOURCE: Infect. Immun. (1999), 67(6), 2941-2950  
 CODEN: INFIBR; ISSN: 0019-9567  
 PUBLISHER: American Society for Microbiology  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English

AB The authors have used expression screening of a genomic Mycobacterium tuberculosis library with tuberculosis (TB) patient sera to identify novel genes that may be used **diagnostically** or in the development of a TB vaccine. Using this strategy, the authors have cloned a novel gene, termed mtb39a, that encodes a 39-kDa protein. Mol. characterization revealed that mtb39a is a member of a family of three highly related genes that are conserved among strains of M. tuberculosis and Mycobacterium bovis BCG but not in other mycobacterial species tested. Immunoblot anal. demonstrated the presence of Mtb39A in M. tuberculosis lysate but not in culture filtrate proteins (CFP), indicating that it is not a secreted antigen. This conclusion is strengthened by the observation that a human T-cell clone specific for purified recombinant Mtb39A protein recognized autologous dendritic cells **infected** with TB or pulsed with purified protein deriv. (PPD) but did not respond to M. tuberculosis CFP. Purified recombinant Mtb39A elicited strong T-cell proliferative and gamma interferon responses in peripheral blood mononuclear cells from 9 of 12 PPD-pos. individuals tested, and overlapping peptides were used to identify a min. of 10 distinct T-cell epitopes. Addnl., mice immunized with mtb39a DNA have shown increased protection from M. tuberculosis challenge, as indicated by a redn. of **bacterial** load. The human T-cell responses and initial animal studies provide support for further evaluation of this antigen as a possible component of a subunit vaccine for M. tuberculosis.

REFERENCE COUNT: 44

REFERENCE(S): (1) Abou Zeid, C; Infect Immun 1988, V56, P3046 HCAPLUS  
 (2) Amara, R; Infect Immun 1996, V64, P3765 HCAPLUS  
 (3) Andersen, P; Infect Immun 1994, V62, P2536 HCAPLUS  
 (4) Andersen, P; J Immunol 1995, V154, P3359 HCAPLUS  
 (8) Barnes, P; J Immunol 1992, V148, P1835 HCAPLUS  
 ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 17 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1998:790680 HCAPLUS  
 DOCUMENT NUMBER: 130:33962  
 TITLE: antigenic peptide and nucleic acid sequences useful for diagnosis of tuberculosis and methods for their use  
 INVENTOR(S): Alderson, Mark R.; Dillon, Davin C.; **Skeiky, Yasir A. W.**; Campos-Neto, Antonio  
 PATENT ASSIGNEE(S): Corixa Corp., USA  
 SOURCE: PCT Int. Appl., 104 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent

**THIS PAGE BLANK (USPTO)**

LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 2  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9853076	A2	19981126	WO 1998-US10514	19980520
WO 9853076	A3	19990401		
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 2001012888	A1	20010809	US 1998-73009	19980505
AU 9875916	A1	19981211	AU 1998-75916	19980520
BR 9809443	A	20000613	BR 1998-9443	19980520
NO 9905690	A	20000118	NO 1999-5690	19991119

## PRIORITY APPLN. INFO.:

US 1997-858998 A 19970520  
 US 1998-73009 A 19980505  
 WO 1998-US10514 W 19980520

AB Claimed are DNA and protein sequences derived from Mycobacterium tuberculosis antigens, which may be used for **diagnosis** of tuberculosis, and methods for such **diagnosis** with immunol. or genetic methods. The compds. provided include polypeptides that contain at least one antigenic portion of one or more M. tuberculosis proteins, and DNA sequences encoding such polypeptides. **Diagnostic** kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of M. tuberculosis **infection** in patients and biol. samples. Antibodies directed against such polypeptides are also provided.

L12 ANSWER 18 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1998:790679 HCAPLUS

DOCUMENT NUMBER: 130:35367

TITLE: Mycobacterium antigenic peptides for immunotherapy and diagnosis of tuberculosis and methods of their use

INVENTOR(S): Alderson, Mark R.; Dillon, Davin C.; **Skeiky, Yasir A. W.**; Campos-Neto, Antonio

PATENT ASSIGNEE(S): Corixa Corp., USA

SOURCE: PCT Int. Appl., 100 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

## PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9853075	A2	19981126	WO 1998-US10407	19980520
WO 9853075	A3	19990401		
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES,				

**THIS PAGE BLANK (USPTO)**

FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI,  
CM, GA, GN, ML, MR, NE, SN, TD, TG

AU 9876907 A1 19981211 AU 1998-76907 19980520  
BR 9809445 A 20000613 BR 1998-9445 19980520  
EP 1012293 A2 20000628 EP 1998-924827 19980520  
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,  
IE, FI

NO 9905689 A 20000118 NO 1999-5689 19991119

PRIORITY APPLN. INFO.: US 1997-859381 A 19970520  
US 1998-73010 A 19980505  
WO 1998-US10407 W 19980520

AB Mycobacterium antigenic peptides, their nucleotide coding sequences, and their use in vaccines and skin testing for tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one immunogenic portion of one or more M. tuberculosis proteins and DNA mols. encoding such polypeptides. Such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against *M. tuberculosis* infection, or may be used for the **diagnosis** of tuberculosis.

L12 ANSWER 19 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1998:550516 HCAPLUS  
DOCUMENT NUMBER: 129:187081  
TITLE: Antigens of Leishmania for use in the prophylaxis, diagnosis, and therapy of leishmaniasis  
INVENTOR(S): Reed, Steven G.; Campos-Neto, Antonio; Webb, John R.; Dillon, Davin C.; **Skeiky, Yasir A.**  
PATENT ASSIGNEE(S): Corixa Corporation, USA  
SOURCE: PCT Int. Appl., 195 pp.  
CODEN: PIXXD2  
DOCUMENT TYPE: Patent  
LANGUAGE: English  
FAMILY ACC. NUM. COUNT: 1  
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9835045	A2	19980813	WO 1998-US3002	19980212
WO 9835045	A3	19981029		
W: BR, MX				
RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
EP 981624	A2	20000301	EP 1998-907489	19980212
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
BR 9807332	A	20001031	BR 1998-7332	19980212
EP 1113073	A2	20010704	EP 2001-101398	19980212
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
PRIORITY APPLN. INFO.:			US 1997-798841	A 19970212
			US 1997-920609	A 19970827
			EP 1998-907489	A3 19980212
			WO 1998-US3002	W 19980212

AB Antigens proteins that can be used in the prevention, treatment and **diagnosis** of leishmaniasis are disclosed. CDNAs for antigens of Leishmania major amastigotes were cloned by screening an expression library with antiserum from **infected** mice. The corresponding antigen was manufd. with a hexahistidine affinity label in Escherichia coli and used to immunize mice. Antisera from these mice reacted with all stages of the parasite lifecycle tested (amastigotes, promastigotes).

**THIS PAGE BLANK (USPTO)**

Antigenic peptides of *L. donovani* were identified by their binding to class II MHC antigens and the gene cloned by PCR using amino acid sequence-derived primers. The protein was manufd. as a fusion protein with glutathione-S-transferase and used to immunize rabbits. Antibody from these rabbits recognized a single protein of 23 kilodaltons that appeared to be presented on the cell surface. The protein also stimulated proliferation of T-cells from mice inoculated with *L. donovani* promastigotes. Use of these proteins to induce protective immunity in a mouse exptl. leishmaniasis model.

L12 ANSWER 20 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1998:251273 HCAPLUS

DOCUMENT NUMBER: 128:307522

TITLE: Antigenic polypeptides of *Mycobacterium tuberculosis* and their encoding DNA sequences for immunotherapy and immunotherapy and diagnosis of tuberculosis

INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**;  
Dillon, Davin C.; Campos-Neto, Antonio; Houghton,  
Raymond; Vedvick, Thomas S.; Twardzik, Daniel R.;  
Lodes, Michael J.

PATENT ASSIGNEE(S): Corixa Corporation, USA

SOURCE: PCT Int. Appl., 230 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9816646	A2	19980423	WO 1997-US18293	19971007
WO 9816646	A3	19981008		
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 6290969	B1	20010918	US 1997-818112	19970313
AU 9748144	A1	19980511	AU 1997-48144	19971007
EP 932681	A2	19990804	EP 1997-910873	19971007
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
CN 1241212	A	20000112	CN 1997-180501	19971007
BR 9712518	A	20001024	BR 1997-12518	19971007
JP 2001501832	T2	20010213	JP 1998-518456	19971007
NO 9901694	A	19990610	NO 1999-1694	19990409
PRIORITY APPLN. INFO.:				
			US 1996-730510	A 19961011
			US 1997-818112	A 19970313
			US 1995-523436	B2 19950901
			US 1995-533634	B2 19950922
			US 1996-620874	B2 19960322
			US 1996-659683	A2 19960605
			US 1996-680574	A2 19960712
			WO 1997-US18293	W 19971007
AB	Antigenic polypeptides that are useful for the immunotherapy and <b>diagnosis</b> of tuberculosis are isolated from the culture filtrate			

**THIS PAGE BLANK (USPTO)**

of *M. tuberculosis* strain H37Ra or H37Rv. The antigen-encoding sequences were also isolated from the genomic DNA libraries of the 2 strains. Such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against *M. tuberculosis* **infection**, or may be used for the **diagnosis** of tuberculosis. **Diagnostic** kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of *M. tuberculosis* **infection** in patients and biol. samples. Antibodies directed against such polypeptides are also provided.

L12 ANSWER 21 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1998:251272 HCAPLUS

DOCUMENT NUMBER: 128:293966

TITLE: Antigenic polypeptides of *Mycobacterium tuberculosis* and their encoding DNA sequences for diagnosis of tuberculosis

INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**;  
Dillon, Davin C.; Campos-Neto, Antonio; Houghton,  
Raymond; Vedvick, Thomas S.; Twardzik, Daniel R.;  
Lodes, Michael J.

PATENT ASSIGNEE(S): Corixa Corporation, USA

SOURCE: PCT Int. Appl., 251 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9816645	A2	19980423	WO 1997-US18214	19971007
W:	AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG			
AU 9747505	A1	19980511	AU 1997-47505	19971007
EP 934415	A2	19990811	EP 1997-910030	19971007
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI			
CN 1242047	A	20000119	CN 1997-180551	19971007
BR 9712298	A	20001024	BR 1997-12298	19971007
JP 2001500383	T2	20010116	JP 1998-518432	19971007
NO 9901693	A	19990609	NO 1999-1693	19990409
PRIORITY APPLN. INFO.:			US 1996-729622	A 19961011
			US 1997-818111	A 19970313
			WO 1997-US18214	W 19971007

AB Antigenic polypeptides that are useful for the **diagnosis** of tuberculosis are isolated from the culture filtrate of *M. tuberculosis* strain H37Ra or H37Rv. The antigen-encoding sequences were also isolated from the genomic DNA libraries of the 2 strains. **Diagnostic** kits contg. such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of *M. tuberculosis* **infection** in patients and biol. samples. Antibodies directed against such polypeptides are also provided.

**THIS PAGE BLANK (USPTO)**

L12 ANSWER 22 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1997:443326 HCAPLUS  
 DOCUMENT NUMBER: 127:64504  
 TITLE: Polypeptide epitopes of TcD, TcE or PEP-2 for  
 detection and prevention of Trypanosoma cruzi  
 infection  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**; Lodes,  
 Michael J.; Houghton, Raymond L.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 110 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 4  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9718475	A1	19970522	WO 1996-US18624	19961114
W: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: KE, LS, MW, SD, SZ, UG, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 5916572	A	19990629	US 1995-557309	19951114
AU 9710568	A1	19970605	AU 1997-10568	19961114
AU 722152	B2	20000720		
EP 874992	A1	19981104	EP 1996-941416	19961114
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
JP 2000503111	T2	20000314	JP 1997-519161	19961114
BR 9611455	A	20010102	BR 1996-11455	19961114
US 6054135	A	20000425	US 1997-834306	19970415
PRIORITY APPLN. INFO.: US 1995-557309 A 19951114				
WO 1996-US18624 W 19961114				

AB Compds. and methods are provided for **diagnosing** Trypanosoma cruzi **infection**. The disclosed compds. are polypeptides contg. TcD, TcE or PEP-2 antigen epitope of Trypanosoma cruzi, or antibodies to the antigens. The polypeptides are coupled to nitrocellulose, latex, or plastic material, and used with labeled binding agent (e.g. anti-Ig., protein G, protein A or lectins) for immunoassay of Trypanosoma cruzi in blood, serum, plasma, saliva, cerebrospinal fluid or urine samples. The polypeptide compds. are further useful in vaccines and pharmaceutical compns. for inducing protective immunity against Chagas' disease in individuals exposed to T. cruzi.

L12 ANSWER 23 OF 27 HCAPLUS COPYRIGHT 2001 ACS  
 ACCESSION NUMBER: 1997:309979 HCAPLUS  
 DOCUMENT NUMBER: 126:276349  
 TITLE: Immunogenic Mycobacterium tuberculosis antigens and  
 genes for immunotherapy and diagnosis of tuberculosis  
 INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**;  
 Dillon, Davin C.; Campos-Neto, Antonio; Houghton,  
 Raymond; Vedvick, Thomas H.; Twardzik, Daniel R.  
 PATENT ASSIGNEE(S): Corixa Corporation, USA  
 SOURCE: PCT Int. Appl., 167 pp.

**THIS PAGE BLANK (USPTO)**

CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 3  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9709428	A2	19970313	WO 1996-US14674	19960830
WO 9709428	A3	19970717		
W: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: KE, LS, MW, SD, SZ, UG, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML				
CA 2230885	AA	19970313	CA 1996-2230885	19960830
ZA 9607394	A	19970505	ZA 1996-7394	19960830
EP 851927	A2	19980708	EP 1996-933009	19960830
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, FI				
CN 1200147	A	19981125	CN 1996-197639	19960830
BR 9610262	A	19990706	BR 1996-10262	19960830
JP 2001517069	T2	20011002	JP 1997-511464	19960830
AU 9671586	A1	19970327	AU 1996-71586	19960930
AU 727602	B2	20001214		
NO 9800883	A	19980427	NO 1998-883	19980227

## PRIORITY APPLN. INFO.:

US 1995-523436	A	19950901
US 1995-533634	A	19950922
US 1996-620874	A	19960322
US 1996-659683	A	19960605
US 1996-680574	A	19960712
WO 1996-US14674	W	19960830

AB Compds. and methods for inducing protective immunity against tuberculosis are disclosed. The compds. provided include polypeptides that contain at least one immunogenic portion of one or more M. tuberculosis antigens and DNA mols. encoding such proteins. Such compds. may be formulated into vaccines and/or pharmaceutical compns. for immunization against M. tuberculosis **infection**, or may be used for the **diagnosis** of tuberculosis.

L12 ANSWER 24 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1997:296911 HCAPLUS

DOCUMENT NUMBER: 126:276348

TITLE: Antigenic polypeptides of Mycobacterium tuberculosis and their encoding DNA sequences for diagnosis of tuberculosis

INVENTOR(S): Reed, Steven G.; **Skeiky, Yasir A. W.**; Dillon, Davin C.; Campos-Neto, Antonio; Houghton, Raymond; Vedvick, Thomas H.; Twardzik, Daniel R.

PATENT ASSIGNEE(S): Corixa Corporation, USA

SOURCE: PCT Int. Appl., 189 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

**THIS PAGE BLANK (USPTO)**

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9709429	A2	19970313	WO 1996-US14675	19960830
WO 9709429	A3	19970717		
W: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: KE, LS, MW, SD, SZ, UG, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML				
CA 2230927	AA	19970313	CA 1996-2230927	19960830
AU 9671587	A1	19970327	AU 1996-71587	19960830
ZA 9607395	A	19970424	ZA 1996-7395	19960830
EP 850305	A2	19980701	EP 1996-933010	19960830
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
CN 1200146	A	19981125	CN 1996-197467	19960830
BR 9610268	A	19990706	BR 1996-10268	19960830
JP 11514217	T2	19991207	JP 1996-511465	19960830
PRIORITY APPLN. INFO.:			US 1995-523435	A 19950901
			US 1995-532136	A 19950922
			US 1996-620280	A 19960322
			US 1996-658800	A 19960605
			US 1996-680573	A 19960712
			WO 1996-US14675	W 19960830
AB Antigenic polypeptides that useful for the diagnosis of tuberculosis are isolated from the culture filtrate of M. tuberculosis strain HH37Ra or H37Rv. The antigen-encoding sequences were also isolated from the genomic DNA libraries of the 2 strains. Diagnostic kits contg. such polypeptides or DNA sequences and a suitable detection reagent for the detection of M.tuberculosis in biol. samples are provided. Antibodies directed against such polypeptides are also provided.				
L12 ANSWER 25 OF 27 HCAPLUS COPYRIGHT 2001 ACS				
ACCESSION NUMBER:		1995:891909 HCAPLUS		
DOCUMENT NUMBER:		124:46736		
TITLE:		Cloning and expression of a Leishmania donovani gene instructed by a peptide isolated from major histocompatibility complex class II molecules of infected macrophages		
AUTHOR(S):		Campos-Neto, Antonio; Soong, Lynn; Cordova, Jose L.; Sant'Angelo, Derek; <b>Skeiky, Yasir A. W.</b> ; Ruddie, Nancy H.; Reed, Steven G.; Janeway, Charles, Jr.; McMahon-Pratt, Diane		
CORPORATE SOURCE:		Medical Sch. Ribeirao Preto, Univ. Sao Paulo, Sao Paulo, Brazil		
SOURCE:		J. Exp. Med. (1995), 182(5), 1423-33 CODEN: JEMEAV; ISSN: 0022-1007		
DOCUMENT TYPE:		Journal		
LANGUAGE:		English		
AB The studies reported here describe the isolation of peptides from MHC class II mols. of murine macrophages <b>infected</b> with Leishmania donovani, and the use of the derived peptide sequences to rescue the pathogen peptide donor protein. The isolation of the peptides was carried out by comparing the RP HPLC profile of peptides extd. from <b>infected</b> macrophages with the peptides extd. from noninfected cells. Several distinct HPLC peaks unique to <b>infected</b> macrophages was sequenced. One of the peptides that was not homologous to				

**THIS PAGE BLANK (USPTO)**

any known protein was used to instruct the designing of an oligonucleotide sense primer that was used in combination with an oligo dT nucleotide (anti-sense primer) to amplify by PCR a DNA fragment from *L. donovani* cDNA. The amplified DNA fragment was cloned and used as a probe to screen a *L. donovani* cDNA library. The cloned gene (Ld peptide gene) has an open reading frame of 525 bp and has no homol. with any known protein/gene sequence. Northern blot analyses indicated that the Ld peptide/gene is broadly distributed and expressed among species of the *Leishmania* genus, in both the amastigote and promastigote life cycle forms. Using the pGEX 2T vector, the gene was expressed and the relationship of the purified recombinant protein with *L. donovani* was confirmed using both antibody and T cell responses from immunized or **infected** animals. The gene encodes a 23-kD mol. (Ldp 23) assocd. with the cell surface of *L. donovani* promastigotes. In addn., T cells purified from the lymph nodes of BALB/c mice immunized with *L. donovani* or **infected** with *L. major*, and from CBA/J mice **infected** with *L. amazonensis* were stimulated to proliferate by the recombinant Ldp 23 and produced high levels of IFN- $\gamma$  and no IL 4. This observation suggests that the Ldp 23 is an interesting parasite mol. for the studies concerning the host-parasite interaction because the Th1 pattern of cytokine response that it induces is correlated with resistance to *Leishmania* **infections**. These results clearly point to an alternative strategy for the purifn. of proteins useful for the development of both vaccines and immunol. **diagnostic** tools not only against leishmaniasis but also for other diseases caused by intracellular pathogens.

L12 ANSWER 26 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1991:580954 HCAPLUS

DOCUMENT NUMBER: 115:180954

TITLE: Immunological characterization of recombinant antigens isolated from a *Mycobacterium avium* .lambda.gt11 expression library by using monoclonal antibody probes  
AUTHOR(S): Rouse, David A.; Morris, Sheldon L.; Karpas, Arthur B.; Mackall, Julia C.; **Probst, Peter G.**;  
Chaparas, Sotiros D.

CORPORATE SOURCE: Cent. Biol. Eval. Res., Food Drug Adm., Bethesda, MD, 20892, USA

SOURCE: Infect. Immun. (1991), 59(8), 2595-600

CODEN: INFIBR; ISSN: 0019-9567

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Nontuberculous mycobacteria, particularly *M. avium*, have been isolated from a significant percentage of patients with AIDS. Early detection of *M. avium* **infection** is difficult, and treatment regimens are often ineffective. Much needs to be learned about antigens and factors responsible for immunity to and pathogenesis of the disease. Specific antigens and **diagnostic** procedures for **infection** need to be developed. To address some of these problems, monoclonal antibodies were generated against a serovar 4 strain of *M. avium* isolated from a patient with AIDS. Protease sensitivity studies have demonstrated that each of these antibodies recognizes a protein-assocd. epitope. Immunoblot analyses suggest that seven of these monoclonal antibodies react specifically with *M. avium* and *M. intracellulare* epitopes. Immunoreactive **bacteriophages** were identified from an *M. avium* .lambda.gt11 expression library with two of these monoclonal antibodies (3808 C3 and 3954 B12). Lambda lysogens, generated from the immunoreactive **bacteriophages**, overproduced .beta.-galactosidase fusion proteins which were reactive with the two monoclonal antibodies in immunoblot assays. The purified fusion proteins were shown to elicit skin test

**THIS PAGE BLANK (USPTO)**

reactions in sensitized guinea pigs.

L12 ANSWER 27 OF 27 HCAPLUS COPYRIGHT 2001 ACS

ACCESSION NUMBER: 1990:457041 HCAPLUS

DOCUMENT NUMBER: 113:57041

TITLE: Production, characterization, and species specificity of monoclonal antibodies to Mycobacterium avium complex protein antigens

AUTHOR(S): Rouse, David A.; Morris, Sheldon L.; Karpas, Arthur B.; Probst, Peter G.; Chaparas, S. D.

CORPORATE SOURCE: Cent. Biol. Eval. Res., Food Drug Adm., Bethesda, MD, 20892, USA

SOURCE: Infect. Immun. (1990), 58(5), 1445-9

CODEN: INFIBR; ISSN: 0019-9567

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The incidence of M. avium-M. intracellulare complex **infections** has increased in recent years primarily because a significant proportion of acquired immunodeficiency syndrome patients develop disseminated M. avium complex disease. In an effort to develop new tools to study these **infections**, 8 monoclonal antibodies directed against M. avium were produced. Western blot (immunoblot) specificity anal. and protease sensitivity assays indicate that 4 of these antibodies recognize M. avium-specific protein epitopes and 2 react with M. avium complex-specific peptide determinants. These monoclonal antibodies may be useful clin. in the **diagnosis** of M. avium complex disease and in the lab. for isolation and characterization of native and recombinant M. avium complex antigens.

**THIS PAGE BLANK (USPTO)**